THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XXII November, 1931 Number 8

# SCHOOL ACHIEVEMENT IN RELATION TO MENTAL AGE—A COMPARATIVE STUDY*

ANDREW W. BROWN AND CHRISTINE LIND

Institute for Juvenile Research (Chicago, Ill.)

## I. Statement of Problem

It is fairly well established that in the regular public school where the children are of average intelligence, the achievement of those with intelligence quotients below one hundred is, in general, above that which would be expected of their mental age, while the achievement of the children with intelligence quotients above one hundred is likely to be below their mental age and the further the intelligence quotient is below or above one hundred the greater the discrepancy between the mental age and achievement. In other words, the accomplishment ratio (AR) which is the ratio between the educational age and the mental age correlates negatively with the intelligence quotient. This is the case in the public school where the majority of the children are of average intelligence.

But is it a general characteristic of all dull children that their achievement be above their mental age and of all bright ones that their achievement be below their mental age, or is this only relative to the group in which the children happen to be placed? To be more specific: It is known that children in the regular public school with intelligence quotients from seventy to ninety, have, in general, achievement ratings above their mental age, but if these same children were in a group where they were relatively bright, would their achievement ratings still be above their mental age or would it fall below? Is the difference between the mental age and achievement age the same for each school subject or do some subjects show a greater variation than others? And again, at what level of intelligence is

* Studies from the Institute for Juvenile Research, Chicago, Series C, 165.

the variation the greatest? It is these problems that this study attempts to investigate.

## II. Method of Study

Two groups of children were selected for study, one group from the academic department of the Lincoln State School for the feeble-minded and one from the Glenwood Manual Training School.

At Lincoln only those children are admitted to the academic department who have intelligence quotients above .50. There were four hundred children in the academic department but only those above grade two were used for this study. Shortly after the children are committed to the institution, they are given the Stanford-Binet examination by the resident psychologist. It is the mental age and intelligence quotient secured by this examination that is used here. If the mental age had been secured more than a month prior to the time of the study, it was brought up to date by multiplying the intelligence quotient by the present chronological age.

Glenwood is a private school for dependent boys. There is no selective factor other than dependency in the selection of the students. The average intelligence rating of the entire school population is around .95. The mental age and intelligence quotient used in this study were secured from the Haggerty Intelligence Examination Delta 2.

The children of both groups were given the following tests of the Stanford-Achievement Series:

1. Paragraph Reading.
2. Word Reading.
3. Arithmetic Reasoning.
4. Arithmetic Computation.
5. Spelling.

In addition to the above, samples of writing were secured from the Lincoln Group and scored according to the Thorndike Writing Scale. One hundred twenty-five of this group were given the Good-enough Drawing Test.[2]

---

[1] Terman, L. M., G. M. Ruch, and T. L. Kelley: "Manual of Directions for Stanford-Achievement Tests." World Book Company.

[2] Goodenough, Florence: "Measurement of Intelligence by Drawings." World Book Company.

All of these tests except writing have been standardized for various age levels, and if there were a perfect relation (assuming the tests are equally well standardized) these age levels for the various abilities should presumably be the same. Therefore, the amount of difference between the achievement ages and the mental age is a measure of the degree of relationship.

## III. Presentation of Data

(*A*) *Retarded Group* (*Lincoln*).—Table I shows the difference for the Lincoln group between the mental age and the ages achieved on the other tests for the three levels of intelligence—fifty to fifty-nine IQ, sixty to sixty-nine IQ, seventy to seventy-nine IQ, for each of the different subjects: Reading paragraphs (RP), reading words (RW), arithmetic reasoning (AR), arithmetic computation (AC), spelling (Sp), writing (Wr), and drawing (Dr). The table shows that in reading paragraphs, the sixty-five children from fifty to fifty-nine IQ made a total of five hundred fifty-seven months above their mental age, and two hundred twenty-nine below their mental age. The average algebraic difference [(557–229) ÷ 65)] of 5.04 months indicates a tendency for this group, in general, to make a rating higher than their mental age in this subject. The eighty-two children with IQ's between sixty and sixty-nine make a total of five hundred eighty months above their mental age and four hundred ninety-seven months below their mental age. On the average, they are 1.01 months above their mental age in reading paragraphs. The thirty-six children with IQ's from seventy to seventy-nine make a total of ninety-four months above and three hundred thirty-two months below their mental age, with an average of 11.5 months below their mental age. The differences between the mental age and the educational age for each of the other subjects are read in the same way. The significant facts of this table are presented in Fig. 1.

Figure 1 shows the variation from the mental age of the various abilities at the different levels of intelligence. The mental age is the standard from which the variation is measured and is indicated by the zero line. The difference in months between the mental age and the ages in the different subjects is represented on the horizontal axis. For example: The children with IQ's from fifty to fifty-nine are, on the average, 5.04 months higher in paragraph reading (RP) than their mental age. The children with IQ's from sixty to sixty-nine are 1.01 months higher, and those with IQ's from seventy to seventy-

TABLE I.—LINCOLN GROUP SHOWING THE DIFFERENCES IN MONTHS BETWEEN THE ACHIEVEMENT AGE IN SCHOOL SUBJECTS AND THE MENTAL AGE FOR THE VARIOUS LEVELS OF INTELLIGENCE

| | Fifty to fifty-nine IQ | | | | Sixty to sixty-nine IQ | | | | Seventy to seventy-nine IQ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Number of cases | Total differences in months | | Average algebraic difference | Number of cases | Total differences in months | | Average algebraic difference | Number of cases | Total differences in months | | Average algebraic difference |
| | | Plus | Minus | | | Plus | Minus | | | Plus | Minus | |
| Reading paragraphs | 65 | 557 | 229 | + 5.04 | 82 | 580 | 497 | + 1.01 | 36 | 94 | 332 | - 6.61 |
| R. W | 65 | 825 | 177 | + 9.97 | 82 | 800 | 228 | + 6.98 | 36 | 168 | 412 | - 6.77 |
| Arithmetic reasoning | 66 | 542 | 283 | + 3.92 | 82 | 396 | 439 | - 0.52 | 38 | 69 | 367 | - 7.85 |
| Arithmetic computation | 66 | 658 | 162 | + 7.51 | 82 | 455 | 308 | + 1.79 | 38 | 46 | 444 | -10.47 |
| Spelling | 66 | 53 | 822 | -11.65 | 82 | 71 | 1551 | -18.05 | 38 | 22 | 1119 | -28.87 |
| Writing | 66 | 580 | 489 | + 1.38 | 82 | 427 | 536 | - 1.33 | 38 | 66 | 596 | -13.95 |
| Drawing | 34 | 300 | 363 | - 1.85 | 58 | 571 | 849 | - 4.79 | 28 | 70 | 602 | -22.21 |

nine are, on the average, 6.61 months below their mental age in this subject.

Tables II, III, and IV give the distribution of the differences in months between the mental ages and the achievement ages in each of the school subjects for each level of intelligence. Most of the cases in the fifty to fifty-nine IQ group lie within a range from positive forty months to negative forty months. Drawing shows the greatest variation.

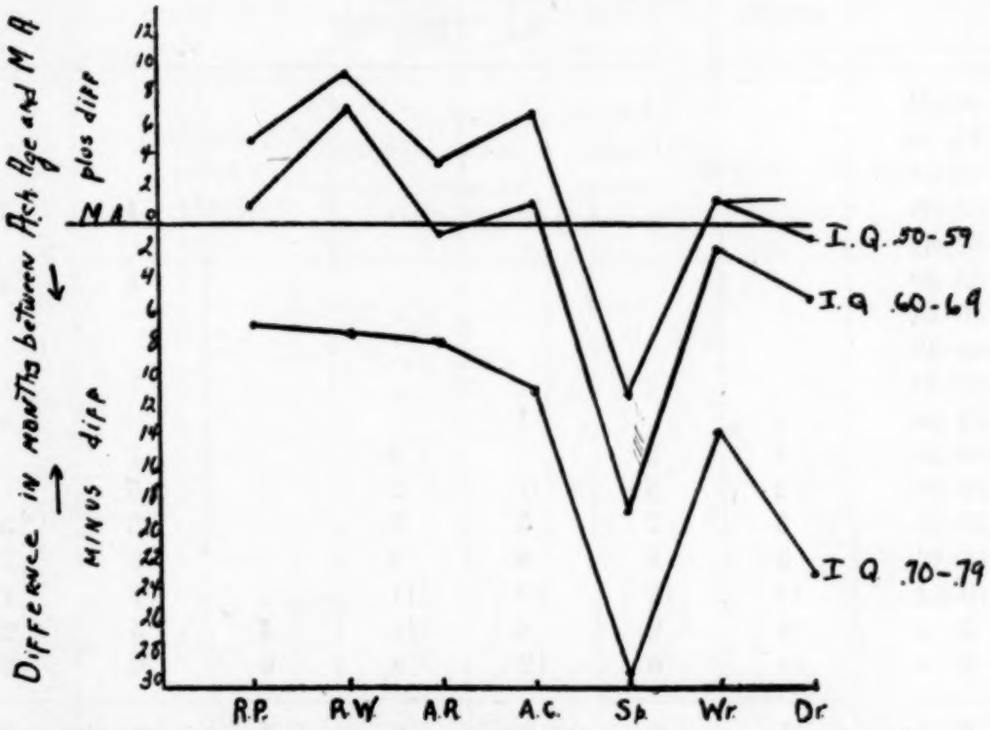



FIG. 1.—Lincoln Group showing the differences in months between the mental age and the achievement ages.

In the sixty to sixty-nine IQ group most of the cases lie within a range from positive thirty months to negative fifty months. In this group also the achievement in drawing has the widest range extending from positive seventy-five months to negative sixty-five months. These two groups have a range of about eighty months. The seventy to seventy-nine IQ group has a smaller range, most of the cases lying between positive twenty months and negative forty months, or a range of sixty months. These tables show that the range of the cases for each subject in the three groups is about the same. In each of the groups, drawing has the widest range. Since all of the ranges are about the same, the distribution of the cases seems to show that each group has about the same scatter and that the significant facts lie

in the change of the average algebraic difference rather than in the range of distribution. The average algebraic differences have been shown in Fig. 1.

TABLE II.—FIFTY TO FIFTY-NINE IQ LINCOLN GROUP SHOWING THE DISTRIBUTION OF DIFFERENCES IN MONTHS BETWEEN MENTAL AGE AND ACHIEVEMENT AGES

| | Reading paragraphs | Reading words | Arithmetic reasoning | Arithmetic computation | Spelling | Writing | Drawing |
|---|---|---|---|---|---|---|---|
| 80–84 | .. | 1 | | | | | |
| 75–79 | | | | | | | |
| 70–74 | | | | | | | |
| 65–69 | .. | .. | .. | .. | .. | 1 | 1 |
| 60–64 | | | | | | | |
| 55–59 | .. | .. | .. | .. | .. | 1 | 1 |
| 50–54 | | | | | | | |
| 45–49 | | | | | | | |
| 40–44 | .. | .. | .. | .. | .. | 1 | |
| 35–39 | 2 | .. | 1 | .. | .. | .. | 1 |
| 30–34 | 1 | 3 | .. | 3 | .. | 1 | |
| 25–29 | 2 | 5 | 1 | 2 | .. | 5 | |
| 20–24 | 5 | 7 | 5 | 7 | .. | 3 | 3 |
| 15–19 | 3 | 8 | 8 | 5 | .. | 2 | 2 |
| 10–14 | 12 | 13 | 12 | 11 | 1 | 7 | 1 |
| 5– 9 | 10 | 9 | 8 | 16 | 4 | 5 | 2 |
| 0– 4 | 11 | 6 | 12 | 8 | 9 | 6 | 3 |
| 5– 1 | 5 | 3 | 6 | 6 | 7 | 6 | 5 |
| 10– 6 | 6 | 5 | 4 | 3 | 10 | 7 | 4 |
| 15–11 | 4 | .. | .. | 1 | 11 | 3 | 2 |
| 20–16 | .. | 2 | 4 | .. | 7 | 9 | 2 |
| 25–21 | 2 | 2 | 1 | 2 | 11 | 7 | 2 |
| 30–26 | 1 | .. | 1 | 1 | 3 | 2 | 1 |
| 35–31 | .. | 1 | 1 | 1 | | | |
| 40–36 | 1 | .. | 2 | .. | 2 | .. | 1 |
| 45–41 | .. | .. | .. | .. | 1 | .. | 1 |
| 50–46 | .. | .. | .. | .. | .. | .. | 1 |
| 55–51 | .. | .. | .. | .. | .. | .. | 1 |

Figure 1 shows that, in general, the lower the intelligence the higher the achievement in relation to the mental age. The children with IQ's between fifty to fifty-nine are doing work decidedly above their mental age level in all the subjects except spelling and drawing.

Those with IQ's between sixty and sixty-nine more nearly approach their mental age in all subjects except spelling and drawing, while those with IQ's from seventy to seventy-nine have achievement levels decidedly below their mental age in all subjects.

(*B*) *Average Group* (*Glenwood*).—The Glenwood Group was selected in order to determine whether the tendency for the brightest retarded children to make achievement ratings below that which would be expected for their mental age, and the dullest ones to have achieve-

TABLE III.—SIXTY TO SIXTY-NINE IQ LINCOLN GROUP SHOWING THE DISTRIBUTION OF DIFFERENCE IN MONTHS BETWEEN MENTAL AGE AND ACHIEVEMENT AGES

| | Reading paragraphs | Reading words | Arithmetic reasoning | Arithmetic computation | Spelling | Writing | Drawing |
|---|---|---|---|---|---|---|---|
| 70–74 | .. | .. | .. | .. | .. | .. | 1 |
| 65–69 | | | | | | | |
| 60–64 | .. | .. | .. | .. | .. | .. | 1 |
| 55–59 | .. | .. | .. | .. | .. | .. | 1 |
| 50–54 | .. | .. | .. | 1 | | | |
| 45–49 | | | | | | | |
| 40–44 | .. | 1 | .. | .. | .. | .. | 1 |
| 35–39 | | | | | | | |
| 30–34 | .. | 1 | .. | 2 | .. | 1 | 2 |
| 25–29 | 3 | 6 | 1 | 1 | .. | 1 | 3 |
| 20–24 | 4 | 6 | 3 | .. | 1 | 4 | 3 |
| 15–19 | 6 | 9 | 7 | 3 | 1 | 6 | 3 |
| 10–14 | 16 | 15 | 6 | 8 | 2 | 5 | 4 |
| 5– 9 | 13 | 11 | 11 | 17 | 1 | 7 | 3 |
| 0– 4 | 9 | 11 | 17 | 16 | 1 | 16 | 7 |
| 5– 1 | 11 | 7 | 10 | 14 | 8 | 11 | 1 |
| 10– 6 | 5 | 7 | 12 | 11 | 9 | 9 | 4 |
| 15–11 | 2 | 4 | 7 | 6 | 7 | 9 | 1 |
| 20–16 | 2 | 1 | 3 | .. | 20 | 4 | 5 |
| 25–21 | 3 | .. | 2 | 2 | 11 | 6 | 1 |
| 30–26 | 3 | 2 | .. | .. | 9 | 2 | 1 |
| 35–31 | 1 | .. | 1 | .. | 3 | 1 | 4 |
| 40–36 | 2 | 1 | 1 | .. | 4 | .. | 4 |
| 45–41 | 1 | .. | .. | .. | 2 | | |
| 50–46 | 1 | .. | .. | .. | 2 | .. | 3 |
| 55–51 | .. | .. | .. | .. | .. | .. | 1 |
| 60–56 | .. | .. | 1 | 1 | 1 | .. | 2 |
| 65–61 | .. | .. | .. | .. | .. | .. | 1 |

ment ratings above their mental age, is typical of the retarded children, or whether the same tendency is present in children who are not retarded.

The Glenwood group represents a fairly average group of boys. The range of intelligence quotients was from fifty-five to about one hundred forty. The average intelligence quotient of the entire school population is about ninety-five.

TABLE IV.—SEVENTY TO SEVENTY-NINE IQ LINCOLN GROUP SHOWING THE DISTRIBUTION OF DIFFERENCE IN MONTHS BETWEEN MENTAL AGE AND ACHIEVEMENT AGES

| | Reading paragraphs | Reading words | Arithmetic reasoning | Arithmetic computation | Spelling | Writing | Drawing |
|---|---|---|---|---|---|---|---|
| 50–54 | | | | | | | |
| 45–49 | .. | 1 | | | | | |
| 40–44 | | | | | | | |
| 35–39 | | | | | | | |
| 30–34 | | | | | | | |
| 25–29 | .. | .. | .. | .. | .. | .. | 1 |
| 20–24 | .. | 1 | .. | .. | 1 | 1 | |
| 15–19 | 2 | 2 | | | | | |
| 10–14 | 1 | 3 | 4 | 1 | .. | 2 | 3 |
| 5– 9 | 5 | 3 | 2 | 4 | .. | 1 | |
| 0– 4 | 4 | 5 | 7 | 4 | .. | 7 | 1 |
| 5– 1 | 7 | 5 | 3 | 4 | 1 | 2 | 2 |
| 10– 6 | 7 | 4 | 8 | 4 | .. | 5 | 4 |
| 15–11 | 3 | 2 | 4 | 9 | 3 | 5 | 1 |
| 20–16 | 1 | 3 | 3 | 4 | 6 | 4 | 2 |
| 25–21 | 2 | .. | 3 | 6 | 5 | 3 | |
| 30–26 | .. | 2 | 2 | 1 | 8 | 2 | 4 |
| 35–31 | 1 | 1 | 1 | .. | 2 | 2 | 2 |
| 40–36 | 2 | 1 | 1 | 1 | 4 | .. | 2 |
| 45–41 | .. | 1 | .. | .. | 2 | 2 | 2 |
| 50–46 | .. | .. | .. | .. | 1 | | |
| 55–51 | .. | .. | .. | .. | 3 | 1 | |
| 60–56 | .. | 2 | .. | .. | 1 | 1 | |
| 65–61 | 1 | .. | .. | .. | 1 | .. | 2 |
| 70–66 | .. | .. | .. | .. | .. | .. | 1 |
| 75–71 | | | | | | | |
| 80–76 | .. | .. | .. | .. | .. | .. | 1 |
| 85–81 | | | | | | | |

The group was divided into six IQ levels: seventy-nine and below, eighty to eighty-nine, ninety to ninety-nine, one hundred to one hundred nine, one hundred ten to one hundred nineteen, one hundred twenty and above. The difference between the mental ages and the achievement ages for the various subjects was computed and tabulated in the same way as for the Lincoln group.

Table V presents the same data for the Glenwood group which Table I presents for the Lincoln group. This table shows that the

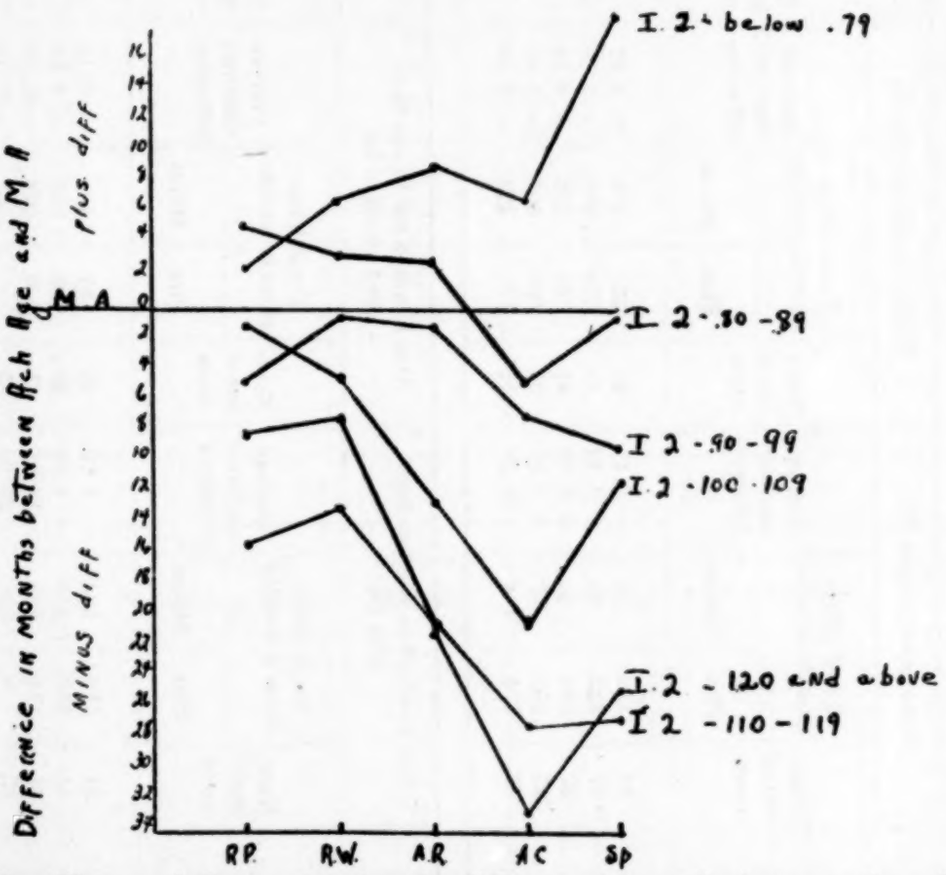



FIG. 2.—Glenwood Group showing the difference in months between the achievement age in subjects and the mental age for the various levels of intelligence.

twenty-seven children with IQ's of seventy-nine and below, made a total of one hundred thirty-eight months above their mental age and eighty-one months below their mental age in reading paragraphs. The algebraic difference ((138 − 81) ÷ 27) of 2.11 months indicates a tendency for this level to make a rating higher than their mental age in this subject. The thirty-one boys with IQ's from eighty to eighty-nine made a total of two hundred fifty-nine months above their mental age and one hundred eleven months below their mental age in reading

TABLE V.—GLENWOOD GROUP SHOWING THE DIFFERENCE IN MONTHS BETWEEN THE ACHIEVEMENT AGE OF SCHOOL SUBJECTS AND THE MENTAL AGE FOR THE VARIOUS LEVELS OF INTELLIGENCE

| | Below seventy-nine IQ | | | | Eighty to eighty-nine IQ | | | | Ninety to ninety-nine IQ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Number of cases | Total differences in months | | Average algebraic difference | Number of cases | Total differences in months | | Average algebraic difference | Number of cases | Total differences in months | | Average algebraic difference |
| | | Plus | Minus | | | Plus | Minus | | | Plus | Minus | |
| Reading paragraphs | 27 | 138 | 81 | + 2.11 | 31 | 259 | 111 | + 4.77 | 41 | 122 | 316 | − 4.73 |
| Reading words | 27 | 232 | 58 | + 6.44 | 31 | 222 | 106 | + 3.74 | 41 | 157 | 173 | − 0.39 |
| Arithmetic reasoning | 27 | 284 | 44 | + 8.88 | 31 | 248 | 138 | + 3.54 | 40 | 172 | 228 | − 1.40 |
| Arithmetic computation | 27 | 253 | 77 | + 6.51 | 31 | 129 | 303 | − 5.61 | 41 | 120 | 412 | − 7.12 |
| Spelling | 27 | 574 | 134 | +16.29 | 31 | 171 | 199 | − 0.90 | 41 | 80 | 449 | − 9.00 |

| | One hundred to one hundred nine IQ | | | | One hundred ten to one hundred nineteen IQ | | | | One hundred twenty and above | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Number of cases | Total differences in months | | Average algebraic difference | Number of cases | Total differences in months | | Average algebraic difference | Number of cases | Total differences in months | | Average algebraic difference |
| | | Plus | Minus | | | Plus | Minus | | | Plus | Minus | |
| Reading paragraphs | 31 | 116 | 192 | − 1.03 | 15 | 43 | 165 | − 8.13 | 16 | 21 | 266 | −15.31 |
| Reading words | 31 | 116 | 252 | − 4.38 | 15 | 23 | 133 | − 7.33 | 16 | 45 | 251 | −12.91 |
| Arithmetic reasoning | 31 | 55 | 450 | −12.61 | 15 | 0 | 303 | −20.20 | 15 | 36 | 341 | −20.33 |
| Arithmetic computations | 30 | 9 | 627 | −20.60 | 15 | | 407 | −27.13 | 16 | 24 | 550 | −32.91 |
| Spelling | 31 | 90 | 457 | −11.86 | 15 | | 406 | −27.06 | 15 | 8 | 389 | −25.06 |

TABLE VI.—SEVENTY-NINE IQ AND BELOW

| | Reading paragraphs | Reading words | Arithmetic reasoning | Arithmetic computation | Spelling |
|---|---|---|---|---|---|
| 40–44 | | | | | |
| 35–39 | | | | | |
| 30–34 | 1 | .. | .. | .. | 1 |
| 25–29 | 1 | 3 | 2 | 1 | 2 |
| 20–24 | .. | 1 | 2 | 2 | 1 |
| 15–19 | 1 | 2 | 5 | 5 | 4 |
| 10–14 | 2 | 3 | 6 | 6 | 1 |
| 5– 9 | 6 | 8 | 5 | 2 | 2 |
| 0– 4 | 3 | 6 | 1 | 2 | 6 |
| 5– 1 | 7 | .. | 3 | 2 | 2 |
| 10– 6 | 3 | 1 | 1 | 4 | 3 |
| 15–11 | 3 | 1 | 1 | 2 | 2 |
| 20–16 | .. | 1 | 1 | 1 | 1 |
| 25–21 | .. | 1 | .. | .. | 1 |
| 30–26 | | | | | |
| 35–31 | | | | | |
| 40–36 | .. | .. | .. | .. | 1 |

TABLE VII.—EIGHTY TO EIGHTY-NINE IQ

| | Reading paragraphs | Reading words | Arithmetic reasoning | Arithmetic computation | Spelling |
|---|---|---|---|---|---|
| 40–44 | | | | | |
| 35–39 | | | | | |
| 30–34 | 1 | 1 | 1 | | |
| 25–29 | 1 | 1 | | | |
| 20–24 | 2 | 3 | 4 | 1 | 2 |
| 15–19 | 4 | 3 | 3 | 2 | 1 |
| 10–14 | 4 | 2 | 3 | 2 | 8 |
| 5– 9 | 4 | 1 | 6 | 5 | |
| 0– 4 | 5 | 8 | 2 | 5 | 6 |
| 5– 1 | 2 | 5 | 2 | 1 | 2 |
| 10– 6 | 3 | 2 | 6 | 3 | 3 |
| 15–11 | 3 | 4 | 1 | 5 | 4 |
| 20–16 | 1 | .. | 1 | 2 | 1 |
| 25–21 | 1 | 1 | 1 | .. | 2 |
| 30–26 | .. | .. | .. | 1 | |
| 35–31 | .. | .. | 1 | 2 | |
| 40–36 | .. | .. | .. | 2 | 1 |

paragraphs. The average algebraic difference for this level was 4.77 months above their mental age. The forty-one boys with IQ's from ninety to ninety-nine had an average of 4.73 months below their mental age in reading paragraphs. The differences between the mental age and subject ages for all the levels are read the same way. The significant facts of this table are presented in Fig. 2; this figure

TABLE VIII.—NINETY TO NINETY-NINE IQ

| | Reading paragraphs | Reading words | Arithmetic reasoning | Arithmetic computation | Spelling |
|---|---|---|---|---|---|
| 40–44 | | | | | |
| 35–39 | .. | .. | .. | 1 | |
| 30–34 | .. | .. | 1 | | |
| 25–29 | | | | | |
| 20–24 | 1 | 2 | .. | 1 | |
| 15–19 | 1 | 2 | 4 | 2 | |
| 10–14 | 4 | 2 | 3 | 1 | 3 |
| 5– 9 | 4 | 8 | 5 | .. | 7 |
| 0– 4 | 1 | 6 | 2 | 5 | 1 |
| 5– 1 | 11 | 7 | 9 | 9 | 8 |
| 10– 6 | 6 | 8 | 7 | 7 | 4 |
| 15–11 | 8 | 3 | 7 | 3 | 4 |
| 20–16 | 1 | 3 | .. | 5 | 5 |
| 25–21 | 1 | .. | .. | 3 | 4 |
| 30–26 | 3 | .. | 1 | 3 | 2 |
| 35–31 | .. | .. | .. | .. | 2 |
| 40–36 | .. | .. | .. | .. | 1 |
| 45–41 | | | | | |
| 50–46 | .. | .. | .. | 1 | |
| 55–51 | | | | | |
| 60–56 | .. | .. | 1 | | |
| 65–61 | | | | | |
| 70–66 | | | | | |
| 75–71 | | | | | |

presents data for the Glenwood group similar to that which Fig. 1 presents for the Lincoln group.

Figure 2 shows the variation from the mental age of the various subjects at the different levels of intelligence. The mental age is the standard from which the variation is measured and is indicated by the zero line. The difference in months between the mental age and

the ages in the different subjects is indicated on the vertical axis. The different subjects are indicated on the horizontal axis.

Tables VI to XI show the distribution of differences in months between the mental ages and the achievement ages in each of the school subjects for each of the IQ levels. In this group most of the cases fall within a range of sixty months. As in the Lincoln group

TABLE IX.—ONE HUNDRED TO ONE HUNDRED NINE IQ

| | Reading paragraphs | Reading words | Arithmetic reasoning | Arithmetic computation | Spelling |
|---|---|---|---|---|---|
| 40–44 | 1 | | | | |
| 35–39 | | | | | |
| 30–34 | 1 | | | | |
| 25–29 | 1 | 1 | | | |
| 20–24 | .. | .. | .. | .. | 1 |
| 15–19 | 2 | 1 | | | |
| 10–14 | 2 | 3 | 1 | .. | 2 |
| 5– 9 | 1 | 4 | 7 | 1 | 2 |
| 0– 4 | 1 | 3 | .. | 2 | 3 |
| 5– 1 | 10 | 4 | 2 | 1 | 2 |
| 10– 6 | 6 | 2 | 3 | 5 | 1 |
| 15–11 | 2 | 6 | 3 | 5 | 5 |
| 20–16 | 2 | 3 | 4 | 3 | 4 |
| 25–21 | .. | 3 | 4 | 3 | 4 |
| 30–26 | 2 | 1 | 3 | 4 | 5 |
| 35–31 | .. | .. | 3 | 3 | 1 |
| 40–36 | .. | .. | 1 | | |
| 45–41 | .. | .. | .. | .. | 1 |
| 50–46 | | | | | |
| 55–51 | .. | .. | .. | 1 | |
| 60–56 | | | | | |
| 65–61 | .. | .. | .. | 1 | |
| 70–66 | | | | | |
| 75–71 | .. | .. | .. | 1 | |

the range of the cases for each subject is about the same. The significant facts seem to lie in the change of the average algebraic differences rather than in the range of distribution.

Tables VI to IX (Glenwood Group) show the distribution of differences in months between mental age and achievement ages for the various IQ levels.

Figure 2 shows the same tendency for this group of boys of average range of intelligence, as Fig. I shows for the retarded group. The brightest group have achievement levels much below that which would be expected of children of their mental age. The dullest group, which included children of the same level of intelligence as the brightest level in the retarded group, make scores on the achievement tests much above their mental age. In general, the lower the intelligence, the higher the achievement in relation to the mental age.

## IV. Discussion of Results

From a comparison of Figs. 1 and 2 it is evident that while the children with IQ's from seventy to seventy-nine in the retarded group are making achievement scores from six to thirty months below their mental age in the different subjects, the children of the same level of intelligence in the average group are making achievement scores from two to sixteen months above their mental age. From this it would appear that the relation of achievement to mental age depends not so much upon the level of intelligence but upon the position of that level in the group receiving instruction.

In studies of children in the public schools, Pintner and others[1] have found this same tendency for the school achievement of bright children to be at a lower level than their mental age, and for the achievement of dull children to be at a higher level than their mental age. That is, the accomplishment ratio (EA/MA) of bright children is, on the average, below 1.00 and for dull children it is above 1.00. Pintner[2] gives as an explanation of this that, "In general, children possessing superior intelligence are the ones who are not working up to possible accomplishment, and the final verdict is that our educational system is failing to make use of the vast stores of intelligence which lie hidden and undiscovered." While it is not true that there are "vast stores" of undiscovered intelligence in the retarded group, it is true that the relatively bright children of this group are not working up to possible accomplishment.

The instruction of the teacher is usually adapted to the ability of the middle group or to the ability of the larger number. If this is so,

---

[1] Pintner, R., and H. Marshall: Results of the Combined-mental Educational Survey Test. *Journal of Educational Psychology*, Vol. XII, No. 2, pp. 82–91.

Forgerson, T. L.: The Efficiency Quotient as a Measure of Achievement. *Journal of Educational Research*, Vol. VI, No. 1, pp. 22–32.

[2] Pintner, R.: "Intelligence Testing," Chapter XI, p. 253.

the duller children have a greater amount of stimulation relative to their ability than the brighter ones, and the tendency would be for those at the extremes to approach the achievement of the average child. This may be one reason why the brighter children are below their mental age in achievement, while the duller ones are above. This,

TABLE X.—ONE HUNDRED TEN TO ONE HUNDRED NINETEEN IQ

| | Reading paragraphs | Reading words | Arithmetic reasoning | Arithmetic computation | Spelling |
|---|---|---|---|---|---|
| 40–44 | | | | | |
| 35–39 | | | | | |
| 30–34 | | | | | |
| 25–29 | | | | | |
| 20–24 | | | | | |
| 15–19 | 2 | | | | |
| 10–14 | 1 | | | | |
| 5– 9 | .. | 2 | | | |
| 0– 4 | 1 | 2 | 1 | | |
| 5– 1 | 2 | 2 | 1 | 1 | |
| 10– 6 | 2 | 2 | 3 | 2 | 1 |
| 15–11 | 3 | 5 | 2 | 2 | 2 |
| 20–16 | 2 | 1 | 1 | 1 | 3 |
| 25–21 | .. | 1 | 1 | 1 | 3 |
| 30–26 | 1 | .. | 3 | 4 | 1 |
| 35–31 | 1 | .. | 1 | .. | 1 |
| 40–36 | .. | .. | 1 | 1 | 1 |
| 45–41 | .. | .. | .. | 1 | |
| 50–46 | .. | .. | .. | .. | 2 |
| 55–51 | | | | | |
| 60–56 | .. | .. | 1 | .. | 1 |
| 65–61 | .. | .. | .. | 2 | |
| 70–66 | | | | | |
| 75–71 | | | | | |

however, should not be so marked in a school where the classes are small.

Again, the tendency has been to keep children of the same chronological age together in spite of differences in ability. The bright children, therefore, are not requird to work up to their possible capacity of achievement, nor are they advanced as far as they should be. The dull children, on the other hand, are urged along in an effort to keep

them in their chronological age group, and, because of this insistent pressure, are likely to acquire training above their capacity.

It is also true that the dull children of any given group have had a longer period of training and experience than the bright children of the

TABLE XI.—ONE HUNDRED TWENTY IQ AND ABOVE

| | Reading paragraphs | Reading words | Arithmetic reasoning | Arithmetic computation | Spelling |
|---|---|---|---|---|---|
| 40–44 | | | | | |
| 35–39 | | | | | |
| 30–34 | | | | | |
| 25–29 | | | | | |
| 20–24 | .. | 1 | 1 | 1 | |
| 15–19 | | | | | |
| 10–14 | 1 | .. | 1 | | |
| 5– 9 | 1 | 3 | .. | .. | 1 |
| 0– 4 | 1 | 2 | | | |
| 5– 1 | 1 | 1 | 1 | .. | 3 |
| 10– 6 | 2 | 1 | 1 | 1 | 3 |
| 15–11 | 3 | 2 | 1 | .. | 1 |
| 20–16 | 3 | 1 | 1 | | |
| 25–21 | 1 | .. | .. | 2 | |
| 30–26 | 1 | .. | 4 | 3 | |
| 35–31 | .. | 1 | 2 | 1 | |
| 40–36 | 1 | 3 | 2 | 4 | 2 |
| 45–41 | .. | .. | 1 | .. | 1 |
| 50–46 | .. | .. | .. | 1 | 2 |
| 55–51 | .. | .. | .. | 1 | 1 |
| 60–56 | .. | 1 | .. | 1 | |
| 65–61 | 1 | .. | .. | 1 | |
| 70–66 | | | | | |
| 75–71 | .. | .. | .. | .. | 1 |

same mental age in the same group. This training would tend to raise the achievement level of the duller children and retard the achievement of the bright ones in relation to their mental age.

Tables X to XI (Glenwood Group) show the distribution of differences in months between mental age and achievement ages for the various IQ levels.

# OBJECTIVE METHODS OF RANKING NURSERY SCHOOL CHILDREN ON CERTAIN ASPECTS OF MUSICAL CAPACITY

THOMAS F. VANCE

Iowa State College

AND

MEDORA B. GRANDPREY

Ohio State University

The elements in the situation which focused attention upon the possiblility of evaluating the musical capacity of nursery school children by a method which might have some approach to reliability were three-fold: an enrollment of thirty-one children, a number of which were almost invariably eager for and interested in the musical phase of the nursery school program; a teacher whose musical ability was considerably above the average, with a professional interest in a problem of this kind; and a psychologist on the staff also interested in the problem.

Certain considerations had, of necessity, to be taken into account in outlining the project. The nursery school at the Iowa State College exists primarily as a laboratory for students in Home Economics. Whatever plan of investigation may be adopted must not interfere with the educational program either of the children themselves or of the college students. The teacher in this instance was employed to teach the children and not to use them as subjects for research.

Out of these two limitations which the situation inherently forces, an important question arises to which the results of this study may give a partial answer; namely, the potential contribution of the nursery school teaching laboratory to the science of child psychology.

The immaturity of nursery school children is an ever present consideration in projecting a work of this character. In fact, were it not for the child's mental immaturity there would have been no point to the project which is being reported. The children would have been rated on their musical capacity according to the Seashore tests. Any method to have any value in determining a child's musical capacity must be exceedingly concrete, must hold his interest and must be within his attention span.

The methods to be described approximate these qualifications. They were concrete, they interested the children and they were of

short duration. Records were obtained on the following capacities or abilities.

1. Responses to music introduced when the children were engaged in other spontaneous interests.
2. Responses to the music played during the regular music period when the children received some encouragement to take part in it.
3. Imitating the nursery school teacher in singing an interval.

(In the instances just cited, the child was responding as a member of the group. In those that follow, the child was alone with the teacher.)

4. Beating time to graphaphone music with the triangle.
5. General responses to music played on the graphaphone.
6. Imitating the nursery school teacher in beating rhythmical patterns on the triangle.
7. Ratings on the basis of the musical aspect of the home environment.

The observations for responsiveness, Tests Nos. 1 and 2, were made on a carefully prepared blank by a student-assistant who knew the children well. They were made two or three days a week for a period of six weeks. A stenographic record was taken for eight different periods. The ratings were based on responses shown through the entire period. A child could be scored as high as five on each of three items, response while playing, presence in circle and response in the circle, with the possibility of a total score of fifteen.

When the graphaphone music was used as a stimulus the teacher simply said to the child, "I have some new records and I thought you would like to hear them." The record was then placed on the machine and the child encouraged to start the machine. The needle was placed in position by the teacher. The teacher then said: "Now, see what you can do to that music." Opportunity for three different responses to each selection was allowed before going to the next record but in each case the only encouragement from the teacher was the question: "What can you do to that music?" For no response the score was 0; for verbal response, 1; movements of parts of the body, 2; movement of the whole body, 3; response appropriate to type of selection, 4.

In the phase of the test where the child played the accompaniment to the graphaphone, a plus was recorded for each measure of four beats each of the march and skip rhythms and for each measure of three beats each of the waltz rhythm played in time to the music. The

skip and march were each sixteen measures in length and the waltz thirty-two measures. The records were Victor 20525-B, 20736-A, and 35774-B.

Best results in the singing tests were obtained when it appeared to the children to be a part of the nursery school program. At the close of the rest period the name of each child was sung, one by one, with the instruction, "You may roll up your rug." To illustrate, let Jimmie be the first child called. "Jimmie, you may roll up your rug." (Jimmie's name was sung by the teacher.) As Jimmie received his direction he would put his rug away and join the teacher at the piano. Another child would then be called, Jimmie following the teacher in singing the name of the second child and so on until all had had their turn. Both ascending and descending intervals were used. Credit was given for responses as follows: Sang one note but off pitch, 1; sang an interval as demonstrated but off pitch, 2; sang an interval of approximate pitch but not on pitch, 3; sang the interval, one note on pitch, 4; sang the interval exactly, 5.

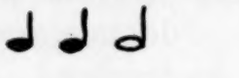
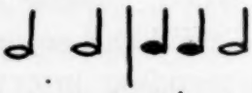
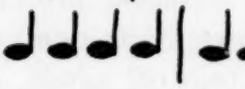
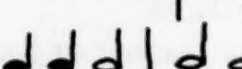
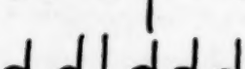
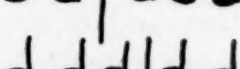
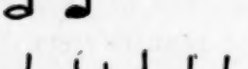
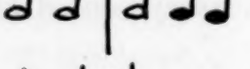
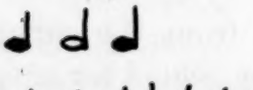
Imitating the nursery school teacher in beating rhythmic patterns on the triangle made the nearest approach to the conventional test situation of any in the series. The patterns themselves, the arrangement, the method of presentation and of scoring which were finally used were all evolved after a rather extended tryout with a three and one-half year old non-nursery school child, a kindergarten child, twenty-one kindergarten children, six kindergarten children and fifty-nine kindergarten children, respectively. There was, however, so little change in the technique of the test after giving it to the fifty-nine kindergarten children that these results may be compared with those obtained from the thirty-one nursery school children who were given the test last of all.

The triangle was selected from the other band and orchestral instruments used in the nursery school for several reasons. It is one of the favorite instruments which was played daily in the nursery school by the children themselves. It was given preference to the xylophone because of its constant pitch; to the cymbals, because it could be manipulated more easily; and to the tambourine, because it offered fewer possibilities for variety in playing.

The method of presenting the problem to the children was as follows: Immediately before the child was brought in to be tested, the teacher checked her tempo with the beat of the metronome which was turned off before the child came in. When the child came in the

teacher began talking about the triangle and showed the child how to hold it. She would grasp it at one of the angles and say, "This is the way I hold it when I play it." Then she would have the child practice taking it, helping him to adjust his fingers to the proper position. In

TABLE I.—ARRANGEMENT OF RHYTHMIC PATTERNS AS GIVEN TO THIRTY-ONE NURSERY SCHOOL CHILDREN WITH TOTAL SCORES FOR EACH PATTERN

| | | |
|---|---|---|
| C2 |  | 49 |
| B3 | | 49 |
| C7 |  | 23 |
| C17 |  | 13 |
| C11 | | 7 |
| C6 |  | 6 |
| C12 |  | 5 |
| C5 |  | 7 |
| C3 | | 7 |
| B2 | | 3 |
| C8 |  | 8 |
| C4 |  | 1 |
| C13 |  | 0 |

playing she was careful to look at the triangle and to strike it in approximately the same place each time. She then said, "I am going to play some little tunes on the triangle. Then you will play them and make them sound just as I do. Listen." Immediately following, the pattern was played and the triangle and bar placed in the child's hands. "Now, make it sound just as I did," was repeated after the

presentation of the first pattern and as frequently afterward as seemed advisable to keep the attention of the child on the problem in hand.

A response was scored as successful when the child reproduced a pattern once, provided the playing showed a difference of long and short notes in the order presented in the pattern, regardless of accent and volume. If played once in conformity to this standard, the score was counted as 1; if played twice, 2; if three times, 3; and if four times, 4. The child was given ample time to play the pattern four times before the record was made on a prepared blank.

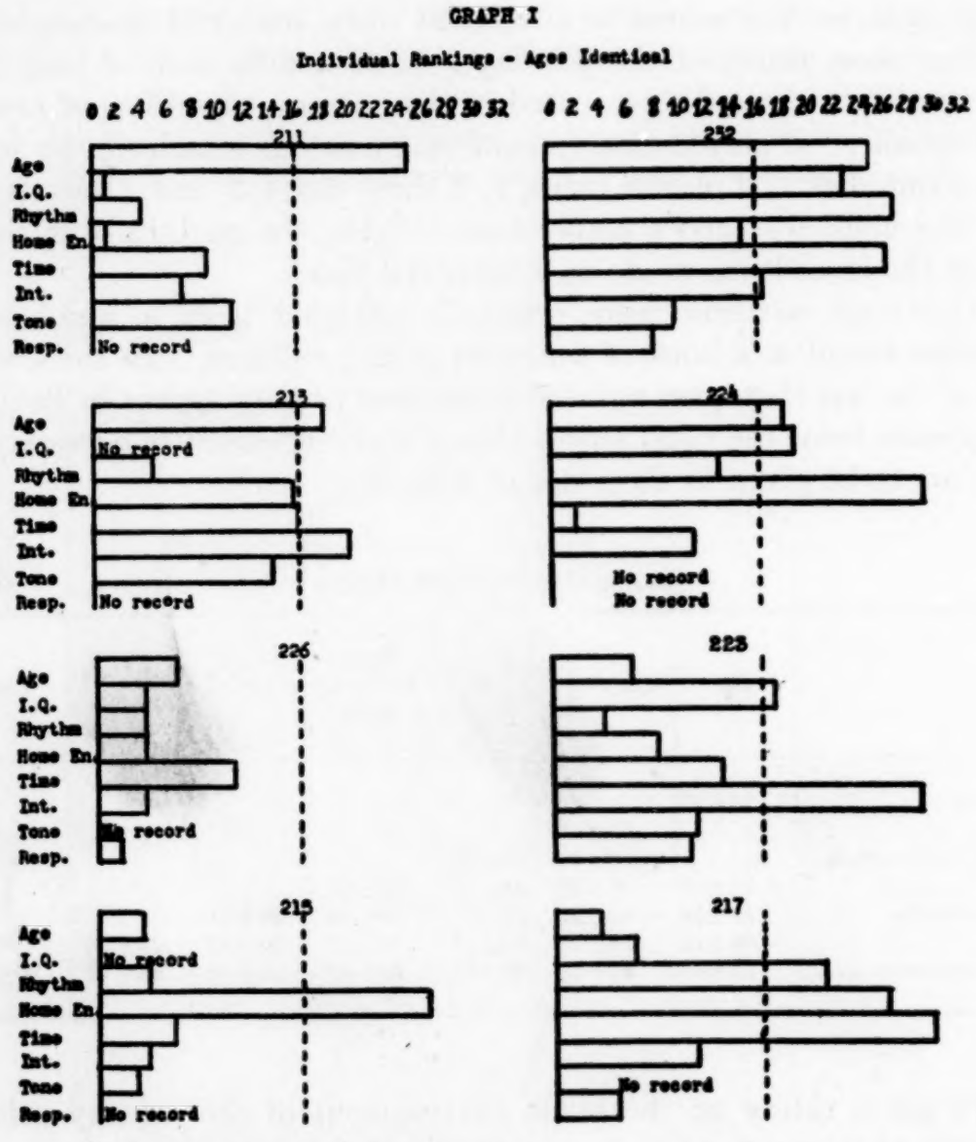
Thirty-six patterns were originally selected from a number of melodies found in a book of songs for young children. In the evolution of the test they were reduced to thirteen as they appear in Table I. It appears from the total scores that a rearrangement is necessary if they are to be given in the order of difficulty.

TABLE II.—CORRELATIONS

| | IQ | Rhythm | Age | Home Environment | Time | Interpretation | Tone |
|---|---|---|---|---|---|---|---|
| Rhythm | .33 ± 12 | | | | | | |
| Age | ..... | .37 ± 11 | | | | | |
| Home Environment | ..... | .28 ± 11 | .05 ± 12 | | | | |
| Time | −.11 ± 12 | .39 ± 10 | .24 ± 11 | .29 ± 11 | | | |
| Interpretation | .12 ± 12 | −.13 ± 12 | .26 ± 11 | .12 ± 12 | .26 ± 11 | | |
| Tone | −.04 ± 15 | .28 ± 13 | .20 ± 14 | .19 ± 14 | .09 ± 15 | .28 ± 13 | |
| Responsiveness | .21 ± 13 | .41 ± 12 | .40 ± 11 | .62 ± 09 | .20 ± 14 | .43 ± 12 | .52 ± 11 |

To get a rating on the home environment of the nursery school children as it pertained to music, the teacher went into the homes of the thirty-one nursery school children with a prepared form which she filled out by the method of direct questioning. The form carried questions relative to the provision for musical impression and expression in the home and the musical training and experience in the home and the musical training and experience of the parents. Scores were given on a basis of from one to five credits on each of the following: instruments in the home, singing, training of parents in music and later experience of parents in music.

The limitations of this paper do not permit giving the results in full. The development of methods of determining special aptitudes




GRAPH I.—Individual rankings—ages identical.

Reading from left to right, individuals of identical ages are grouped in pairs. In the case of the first pair, No. 232 is shown to possess musical capacities of above average or nearly average, while No. 211 falls far below average. No. 223 ranks well above average in time keeping ability, and although he fails to reach the average mark in other capacities, scores well above No. 226 of identical age. A similar comparison may be made in the case of No. 217 and No. 215. Home environment, interpretation and responsiveness to music are abbreviated in the graph respectively as follows: Home En., Int., and Resp.

of young children is at present more important than results. The tables and graphs which accompany may be taken as samples.

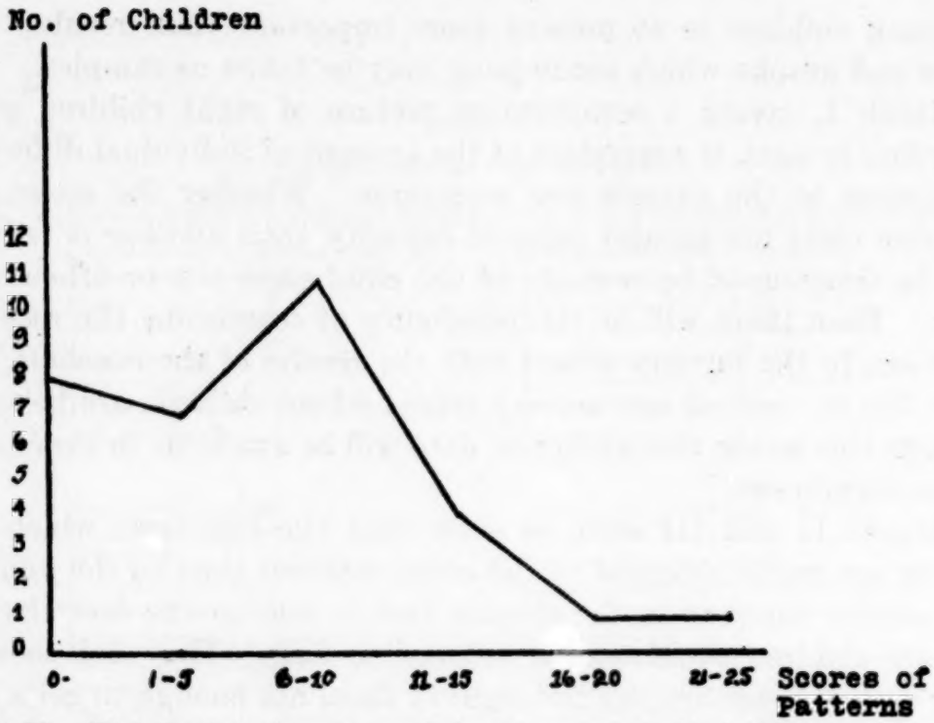
Graph I, giving a comparative picture of eight children paired according to ages, is suggestive of the amount of individual differences in reaction to the various test situations. Whether the assumption that one child has greater musical capacity than another is true can only be determined by a study of the child some ten or fifteen years later. Then there will be the possibility of correlating the results of the work in the nursery school with the results of the Seashore tests. With this in view all our nursery school school children are being put through this series, that sufficient data will be available to make correlations significant.

Graphs II and III seem to show that the two tests which they present are better adapted to the older children then to the younger. The smaller number in the singing test is due to the fact that the younger children could not be induced to sing. They did do something with the rhythm test but eight of them not enough to get a score of more than zero. The tests are not recommended for children younger than three years. In the rhythm test the median for the kindergarten children was 14.8; for the four to five year level, 8.84; the three to four year level, 4.18 and for the two to three year level, 2.43. Further, the correlation of rhythm with age in the kindergarten was .03 while in the nursery school it went as high as .37. To give any test of special aptitude too early may give merely a measure of maturity and not of the aptitude in question.

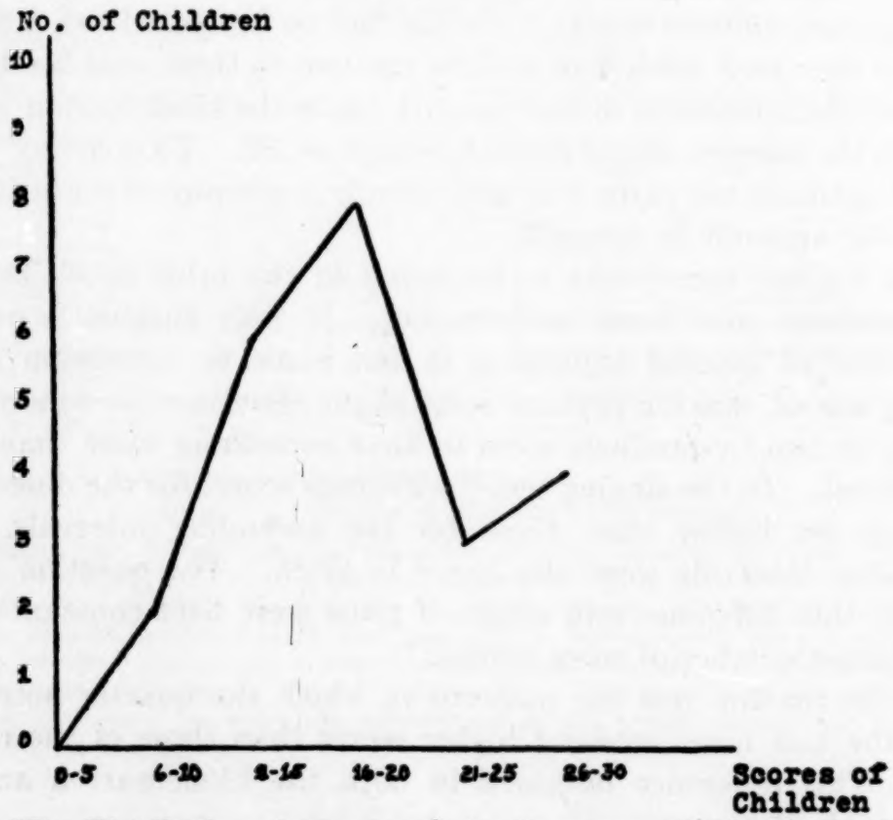
The highest correlation to be found in the table is .62 between responsiveness and home environment. It may suggest a possible carry over of musical expression in the home to expression in the nursery school, making perhaps some slight allowance for inheritance.

One or two by-products seem to have something more than passing interest. In the singing test the average scores for the descending intervals are higher than those for the ascending intervals. The descending intervals were also lower in pitch. The question arises, "Would this difference still obtain if pitch were held constant and a much larger number of cases studied?"

In the rhythm test the patterns in which the quarter notes preceded the half notes received higher scores than those of the reverse order. This difference obtained in both the kindergarten and the nursery school groups.




GRAPH II.—Distribution of scores made by 31 nursery school children on rhythmic pattern test.




GRAPH III.—Distribution of scores made by 23 nursery school children in interval singing test.

Taken as a whole, the study may be used as an illustration of the use of the nursery school as a teaching laboratory, as differentiated from a research laboratory, for securing data which may prove of scientific as well as of practical value. Practical results certainly accrue in the stimulation of the interest of the nursery school staff to secure as reliable information as possible about the children. Only the future can determine the scientific value of such a study.

# FACTUAL MEMORY OF SECONDARY SCHOOL PUPILS FOR A SHORT ARTICLE WHICH THEY READ A SINGLE TIME[1]

ALFRED G. DIETZE AND GEORGE ELLIS JONES

University of Pittsburgh

## Introduction

*The Problem.*—A review of the literature reveals that but few of the many studies of memory touch upon the subject from the point of view of what is actually remembered of things experienced in ordinary life situations. The writers feel that research in this general field should give profitable returns and have undertaken an investigation of a typical memory situation,—namely, that of a single reading of a short informative article by pupils in the secondary schools. The present study is an attempt to determine the range and average extent of immediate and delayed factual memory of pupils in Grades VII to XII for a short article which they read a single time and the relation which obtains between memory at intervals of increasing length.

This and related problems emerge in a consideration of the practical significance of the classical laboratory studies of memory. It is uncertain, for example, whether we remember what we obtain by reading and study in the same way that we remember material learned by rote to the degree of one errorless and unhesitating recital. The pedagogical significance of exact information upon this point relative to a single reading of an article is clear in view of the fact that pupils in school and college commonly prepare assignments by reading lessons a single time.[2]

*Related Studies.*—A number of related investigations have been reported during the past ten years. Jones studied the memory of college students for the content of lectures in psychology.[3] The

---

[1] Acknowledgment is made to Professor H. B. Reed for fundamental suggestions.

[2] *Cf.* Yoakam, G. A.: "Reading and Study." The Macmillan Co., New York, 1928, pp. 185–190.

[3] Jones, H. E.: Experimental Studies of College Teaching. *Archives of Psychology*, No. 68 (1923).

students knew an average of 62 per cent of the facts presented when they were tested immediately after the lectures. After three or four days they remembered an average of 45 per cent, and after eight weeks memory dropped to 24 per cent.

Several writers report conflicting results regarding memory over the summer vacation.[1] Recently Bassett published a study of the retention of history in the sixth, seventh and eighth grades.[2] Objective tests were given at the end of the course and again after intervals of four, eight, twelve and sixteen months. Computing retention on the basis of the scores in the initial tests, average memory scores of 86.01, 81.51, 76.72 and 71.96 per cent respectively were obtained. Since, however, no measures of validity and reliability are presented, and since the tests are short and include four different types of questions to which equal credit was given in the scores, the validity of the conclusions may be questioned. At least the method hardly justifies a statement such as, "It is worthy of note that after sixteen months the children know seventy-two per cent of the history which they knew at the end of the semester."[3]

In respect to method and results there are many points of similarity between Yoakam's study of the effects of a single reading and the present investigation. His finding that the effect of a single reading gives on the average less than half of the total ideas in the article[4] is of especial interest here. He also found that that the effect of a single reading varies with the individual and from grade to grade; and he

---

[1] Garfinkle, M. A.: The Effect of the Summer Vacation on Ability in the Fundamentals of Arithmetic. *Journal of Educational Psychology*, Vol. X, 1919, pp. 44–47; Kirby, T. J.: "Practice in the Case of School Children." Columbia University Contributions to Education, No. 58. Bureau of Publications, Teachers College, Columbia University, New York, 1913; Noonan, M. E.: "Influence of the Summer Vacation on the Abilities of Fifth and Sixth Grade Children." Columbia University Contributions to Education, No. 204. Bureau of Publications, Teachers College, Columbia University, New York, 1926.

[2] Bassett, S. J.: "Retention of History in the Sixth, Seventh and Eighth Grades with Special Reference to the Factors that Influence Retention." The Johns Hopkins University Studies in Education, No. 12. Johns Hopkins Press, Baltimore, 1928.

[3] See *ibid.*, p. 51.

[4] Yoakam, G. A.: The Effects of a Single Reading. *University of Iowa Studies in Education*, Vol. II, No. 7. The University of Iowa, Iowa City, 1924. See especially conclusions on page 98ff.

questions whether a single reading without any immediate recall will leave any impression on the mind of the learner after a lapse of twenty or thirty days, unless the material is very interesting or striking.

## Materials and Method

*Subjects.*—Children in Grades VII to XII acted as subjects, a total of 2789 taking part. The experimental groups consisted of 2062 pupils in the junior and senior high schools of Uniontown, Pa. Each grade was divided into six numerically approximately equal groups for purposes of rotation in the experiments. The remaining seven hundred twenty-seven subject, drawn from neighboring schools, were used in control experiments: Control Group A, two hundred seventy-eight children, for obtaining intercorrelations between the tests; Control Group B, two hundred forty-five pupils, for an investigation of previous knowledge of the articles; Control Group C, one hundred eighty-two pupils, for obtaining a measure of how well pupils comprehend the materials. The median scores of these groups in intelligence tests, reading and vocabulary tests correspond closely to published norms; hence, it may be safely assumed that they are fairly typical of the secondary school population.[1]

*The Reading Selections.*—Three interesting and highly factual articles were used as reading selections, each printed in the form of a four-page pamphlet. "Radium: The Magic Metal," containing 1265 words, is an account of the discovery and uses of radium. "The Early Germans" describes the habits and customs of the early inhabitants of central Europe and is 1061 words in length. "Sir Richard Arkwright" is an account of the life of the English inventor and contains 1279 words. These articles are well suited to the purpose for which they were selected, since they abound in facts upon which questions can be formulated and are of suitable difficulty for the grades studied.[2]

---

[1] Complete tables for all summarizations contained in this paper are to be found in A. G. Dietze, "Factual Memory of Secondary School Pupils for a Short Article Which They Read a Single Time." Doctor's dissertation, Library of the University of Pittsburgh, 1930.

[2] For methods of determining difficulty *cf. ibid.*, pp. 51–56.

*The Tests.*—The tests were made up of five-response multiple choice type questions and were printed in the form of four-page pamphlets. An effort was made to thoroughly canvass the articles for all thought units contained so as to make the memory tests as complete as possible. The test on Radium contains one hundred items, the other two one hundred twenty-five items each.

In respect to validity, objectivity and reliability these tests compare favorably with some of the better tests in standard use. Validity is attained by thoroughly covering the content of the articles in the questions and is rendered evident by an average intercorrelation, corrected for attenuation, of .80 ± .014 between the tests. Objectivity is attained by (*a*) careful instruction for administration, (*b*) clear directions to pupils, (*c*) unambiguous questions, (*d*) uniform and well marked scoring keys in the form of strip stencils, and (*e*) supervision of scoring by one responsible person. The reliability of the tests is shown by the following figures, the reliability coefficients being obtained by applying the Spearman-Brown prophecy formula to the correlation between odd and even halves of the tests:

| | Radium | Germans | Arkwright |
|---|---|---|---|
| Reliability of half test | .82 ± .008 | .89 ± .005 | .85 ± .008 |
| Reliability coefficient | .90 ± .005 | .94 ± .003 | .92 ± .004 |
| Index of reliability | .95 ± .002 | .97 ± .001 | .96 ± .002 |
| PE of estimate | 3.63 | 3.60 | 3.72 |
| PE of measurement | 2.63 | 2.59 | 2.68 |
| Number of cases | 659 | 667 | 616 |

*Scoring.*—Since the tests are designed to measure the amount retained, the best type of score seems to be the percentage of items answered correctly. In this way each item is considered a unit, relative difficulty of the separate questions being disregarded. No attempt is made to correct for guessing, since instructions were given not to guess. It may be stated that this method of scoring resulted in distributions approximating the normal about as well as can be expected from the number used. Arkwright gave an especially good fit to the normal curve.

*Procedure.*—The method of conducting the experiments was to give the pupils an article, with instructions to read it a single time in order to answer a memory test at some time later. The pupil was allowed to read at his own rate. The test was given immediately after the reading in Experiment 1, and after intervals of 1, 14, 30 and 100 days in Experiments 2, 3, 4 and 5 respectively. Each group participated in the immediate memory experiment and in two delayed memory experiments, using different materials for each. The manner in which the groups and materials were rotated appears in the following scheme:

| Group | Experiments participated in and material used in each | | |
|---|---|---|---|
| I | Exp. 1, Radium | Exp. 2, Germans | Exp. 5, Arkwright |
| II | Exp. 1, Radium | Exp. 3, Germans | Exp. 4, Arkwright |
| III | Exp. 1, Germans | Exp. 2, Arkwright | Exp. 5, Radium |
| IV | Exp. 1, Germans | Exp. 3, Arkwright | Exp. 4, Radium |
| V | Exp. 1, Arkwright | Exp. 2, Radium | Exp. 5, Germans |
| VI | Exp. 1, Arkwright | Exp. 3, Radium | Exp. 4, Germans |

The reading and testing periods were administered by the English teachers of the regular classes, thus providing a natural school-room situation to the subjects. The teachers were carefully coached in the manner of conducting the experiments and instructed to report all irregularities to the experimenter. Before the experiments proper were begun, all groups were given a practice exercise consisting of a reading selection and test similar in nature to those above described. This served to acquaint the subjects thoroughly with the details of taking the tests.

## Supplementary Results

Several factors which cannot be eliminated by our method influence the results obtained. The approximate extent to which two of these enter was made the object of two control experiments. The points investigated are (*a*) the influence of previous knowledge and ability to answer questions by logical thinking and (*b*) the failure of pupils to

answer questions because of comprehension difficulties and careless work habits.

*Previous Knowledge.*—Control Group B was required to answer the tests without previously having read the articles. They were, furthermore, instructed to guess when they did not know the answer to a question. The results are somewhat confusing, since the medians are in inverse relation to grade. The reason for this, however, becomes apparent upon examination of the papers: The lower grades complied more fully with the directions regarding guessing and, therefore, had more opportunities of getting items correct by chance. The actual medians range from 32.3 per cent in the seventh grade to 23.5 in the twelfth for Radium, from 30.8 in the seventh to 25.0 in the eleventh for Germans, and from 24.8 in the seventh to 18.5 in the twelfth for Arkwright.

The instructions regarding guessing were given for the purpose of extracting the correctly guessed items from the scores by the commonly used formula for correcting multiple choice tests;[1] however the correction could be accomplished only in the seventh grade, since in this grade the directions were strictly followed. When the scores for the seventh grade are subjected to the formula, the following results remain as the approximate scores that pupils in this grade are able to make without previously reading the articles: Radium 15.4 per cent; Germans 13.4 per cent; Arkwright 6.0 per cent. These amounts are interpreted as due to previous knowledge and logical thinking and are in harmony with the findings of Yoakam in a similar experiment.[2]

*Comprehension Difficulty and Carelessness.*—The ability of pupils to answer the tests when they are allowed to look up the answers in the articles may be taken as a measure of how well they comprehend the materials and how careful they are in their work. Control Group C was used for determining this factor. The means for the respective grades range from 82.0 per cent in the eleventh grade to 66.6 in the seventh for Germans, and from 73.0 in the twelfth to 54.3 in the seventh for Arkwright. The difference between grades are for the most part reliable. The average of all grades in Germans is 75.7, in Arkwright 63.5. Thus this grade missed about twenty-five per cent of the

---

[1] Corrected score = No. right $- \frac{\text{No. wrong}}{(n-1)}$.

[2] *Op. cit.*, pp. 58–59.

questions on Germans and about thirty-seven per cent of those on Arkwright because they did not comprehend the materials read or were too careless to do thoroughly accurate work.

Although the influence of the above factors have not been determined with exactness, nevertheless the control experiments show that it is significant; and the experimental results must, accordingly, be interpreted by the reader with these facts in mind.

## Factual Memory after Five Intervals

*Statistical Treatment.*—Two methods of summarizing the results are employed. The first makes use of the common statistical device of describing group results in terms of means and standard deviations. This method brings out the trend of the data in a straightforward manner which needs no further comment.

The other method is made necessary because the technique of equated parallel groups was not followed. For this reason the various sub-groups are only roughly comparable, the group average depending to a large extent upon the ability of the particular group relative to entire grades or all groups taken together. To obtain comparable results for grades or other sub-groups, it is necessary first to transmute scores derived from the three different materials into equivalent terms and then to combine the results into an average representative of the combined groups.

The best method for accomplishing this is to make use of the regression equations derived from the correlations between the materials. It will be remembered that Control Group A was used for the purpose of getting correlations between the three materials. From these correlations the regression equation between Radium and Arkwright and between Germans and Arkwright were derived, the resulting equations making possible the conversion of results in Radium and Germans into scores equivalent to Arkwright scores and the combination of all three groups in terms of an average for the latter material. To distinguish this derived score from the scores of groups actually using Arkwright the term "A-score" will be used. The actual equations for deriving A-scores from the results in Radium and German were found to be:

$$\text{A-score} = .80\ \text{Radium} + 2.4 = .67\ \text{Germans} + 11.0.$$

Similar scores may, of course, be derived in terms equivalent to Radium or Germans. The reader may get an approximate value for R-scores or G-scores by adding ten points to the A-score averages herein reported.

*Immediate Memory.*—Table I gives the means and standard deviations of the respective groups for the material which they read and the A-score averages of the combined groups in immediate memory.

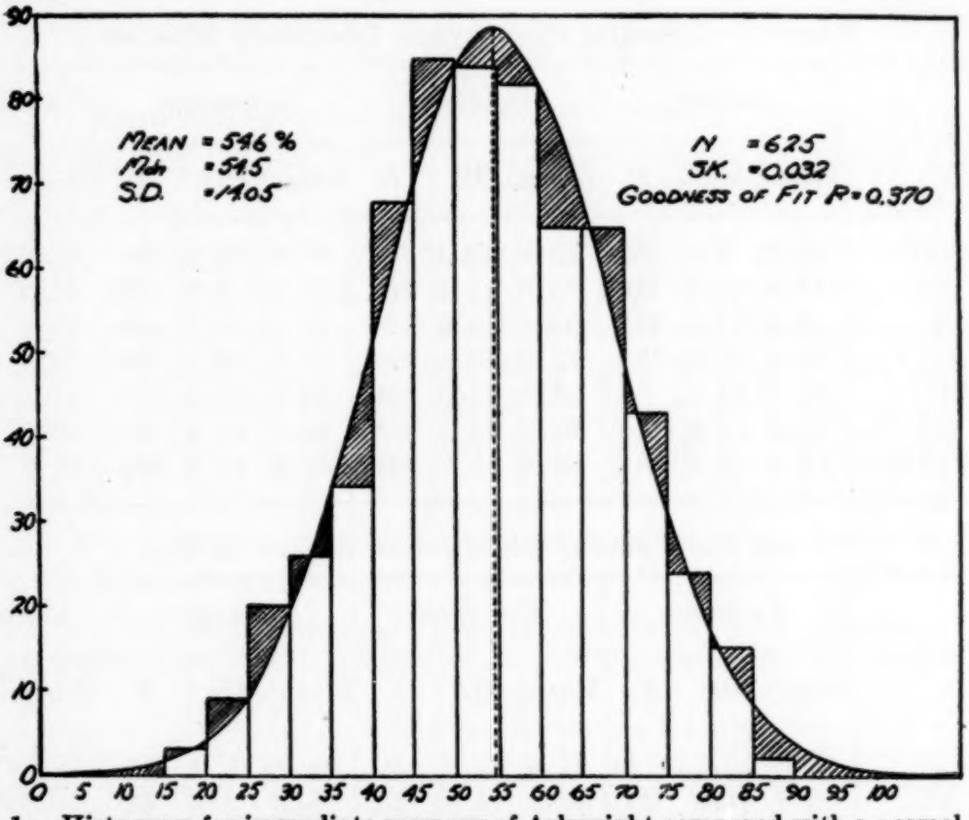



FIG. 1.—Histogram for immediate memory of Arkwright compared with a normal curve fitted to the data.

The total range of the distributions are from thirty-one to one hundred per cent for Radium, from sixteen to one hundred per cent for Germans, and from sixteen to ninety per cent for Arkwright. Thus pupils in the junior and senior high school score over four-fifths of the entire scoring range of these tests, which shows that individual differences in this ability are great. The distributions for Radium and Germans are somewhat negatively skewed, *Sk* being −.352 and −.265 respectively. [1]

[1] $Sk = \frac{3(\text{Mean} - \text{Median})}{\text{SD}}$.

This indicates that there is some piling up of immediate memory scores in the upper part of the range for these articles. Arkwright, however, gives a very good approximation to the Gaussian curve, *Sk* being only .032 and Pearson's chi-square test of goodness of fit giving a *P* value of .370, which is very satisfactory for a population of slightly more than 600.[1] The comparison of the latter distribution with the theoretical normal curve fitted to the data is shown graphically in Fig. 1.

TABLE I.—AVERAGE PERCENTAGE IMMEDIATE MEMORY

| Grade | Radium | | | Germans | | | Arkwright | | | A-Score | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | *N* | Mean | SD | *N* | Mean | SD | *N* | Mean | *N* |
| XII | 84.3 | 7.6 | 68 | 75.8 | 12.5 | 67 | 64.8 | 11.2 | 109 | 65.2 | 244 |
| XI | 77.0 | 10.1 | 116 | 73.0 | 12.2 | 101 | 62.7 | 11.2 | 61 | 62.3 | 278 |
| X | 76.6 | 11.6 | 121 | 66.6 | 12.3 | 122 | 57.4 | 11.9 | 123 | 59.0 | 366 |
| IX | 72.4 | 10.6 | 139 | 67.3 | 13.6 | 109 | 50.3 | 12.1 | 126 | 55.5 | 374 |
| VIII | 66.6 | 12.1 | 72 | 57.9 | 13.5 | 120 | 50.0 | 14.1 | 131 | 51.3 | 323 |
| VII | 67.3 | 12.6 | 146 | 53.8 | 15.6 | 138 | 43.9 | 12.9 | 75 | 50.0 | 359 |
| VII–XII | 73.4 | 12.4 | 662 | 64.8 | 15.7 | 657 | 54.6 | 14.1 | 625 | 56.7 | 1944 |

TABLE II.—PERCENTAGE MEMORY AFTER ONE DAY

| Grade | Radiums | | | Germans | | | Arkwright | | | A-Score | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | *N* | Mean | SD | *N* | Mean | SD | *N* | Mean | *N* |
| XII | 72.9 | 10.7 | 74 | 71.8 | 9.6 | 30 | 63.3 | 8.9 | 35 | 61.1 | 139 |
| XI | 64.9 | 11.8 | 26 | 59.3 | 14.9 | 53 | 55.2 | 9.4 | 52 | 53.2 | 131 |
| X | 56.3 | 10.7 | 43 | 58.8 | 15.7 | 60 | 49.7 | 12.7 | 57 | 47.9 | 160 |
| IX | 56.1 | 13.9 | 62 | 54.2 | 14.8 | 66 | 42.4 | 12.8 | 31 | 46.0 | 159 |
| VIII | 55.0 | 12.3 | 61 | 55.1 | 13.0 | 34 | 43.4 | 10.3 | 57 | 45.3 | 152 |
| VII | 47.7 | 15.4 | 38 | 49.0 | 16.3 | 69 | 33.5 | 8.3 | 71 | 40.0 | 178 |
| VII–XII | 58.7 | 15.1 | 304 | 56.7 | 17.2 | 312 | 46.4 | 15.0 | 303 | 48.0 | 919 |

*Delayed Memory, One Day.*—The average results for memory after one day are given in Table II. The distributions are similar in nature as in the previous experiment. The range in Radium is from 11 to 95 per cent, in Germans from 16 to 95, and in Arkwright from 6 to 85.

[1] See K. J. Holzinger, "Statistical Methods for Students of Education." Ginn and Co., Boston, 1928, pp. 245–248.

Thus the same wide range of individual differences is to be noted as in immediate memory.

*Delayed Memory, Fourteen Days.*—Results for memory after fourteen days are shown in Table III. The distributions at this interval have become positively skewed, showing that the task is becoming more difficult. *Sk* is .191 for Radium, .405 for Germans and .405 for Arkwright. The range in Radium is from eleven to seventy per cent, that in Germans is from six to eighty-five per cent, and that in Arkwright from one to seventy per cent. Again we find ability in these grades to range from very low to very high.

TABLE III.—PERCENTAGE MEMORY AFTER FOURTEEN DAYS

| Grade | Radium | | | Germans | | | Arkwright | | | A-Score | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | *N* | Mean | SD | *N* | Mean | SD | *N* | Mean | *N* |
| XII | 40.1 | 11.8 | 34 | 51.3 | 13.2 | 27 | 32.0 | 10.0 | 27 | 36.8 | 88 |
| XI | 50.6 | 11.4 | 31 | 46.7 | 13.8 | 54 | 32.3 | 12.3 | 42 | 39.0 | 127 |
| X | 40.8 | 10.9 | 74 | 44.2 | 17.6 | 48 | 34.4 | 12.1 | 56 | 36.1 | 178 |
| IX | 44.9 | 9.7 | 64 | 38.8 | 11.0 | 62 | 32.9 | 10.9 | 69 | 35.9 | 195 |
| VIII | 45.2 | 10.6 | 64 | 34.3 | 9.9 | 27 | 28.8 | 11.3 | 56 | 33.8 | 147 |
| VII | 40.7 | 9.5 | 35 | 37.9 | 10.9 | 71 | 37.5 | 15.2 | 59 | 36.3 | 165 |
| VII–XII | 43.7 | 11.0 | 302 | 42.0 | 14.1 | 289 | 33.1 | 12.0 | 309 | 36.3 | 900 |

TABLE IV.—PERCENTAGE MEMORY AFTER THIRTY DAYS

| Grade | Radium | | | Germans | | | Arkwright | | | A-Score | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | *N* | Mean | SD | *N* | Mean | SD | *N* | Mean | *N* |
| XII | 37.9 | 9.8 | 23 | 39.3 | 9.3 | 35 | 31.3 | 11.6 | 30 | 33.8 | 88 |
| XI | 42.1 | 12.6 | 50 | 40.1 | 11.0 | 26 | 31.5 | 12.5 | 48 | 34.9 | 124 |
| X | 42.2 | 11.1 | 54 | 32.1 | 8.5 | 34 | 37.8 | 14.4 | 47 | 35.7 | 135 |
| IX | 37.1 | 11.5 | 65 | 32.1 | 9.1 | 53 | 25.4 | 8.9 | 67 | 29.5 | 185 |
| VIII | 37.7 | 11.1 | 58 | 41.3 | 11.4 | 61 | 27.6 | 7.7 | 26 | 33.2 | 145 |
| VII | 37.2 | 11.1 | 56 | 32.9 | 9.5 | 34 | 25.9 | 9.0 | 68 | 25.5 | 158 |
| VII–XII | 38.3 | 11.6 | 306 | 36.4 | 10.7 | 243 | 30.3 | 12.0 | 286 | 32.4 | 835 |

*Delayed Memory, Thirty Days.*—Average results for memory after thirty days are shown in Table IV. The nature of the distributions (skewness) is the same as in the preceeding experiment. Scores in Radium range from six to eighty per cent, in Germans from six to seventy-five, and in Arkwright from one to seventy-five.

*Delayed Memory, One Hundred Days.*—The results for the fifth experiment are shown in Table V. The scores in Radium range from one to seventy-five per cent, in Germans from one to sixty, and in Arkwright from one to ninety. As in the previous experiment, skewness is slightly positive, showing massing of scores below the mean.

TABLE V.—PERCENTAGE MEMORY AFTER ONE HUNDRED DAYS

| Grade | Radium | | | Germans | | | Arkwright | | | A-Score | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | *N* | Mean | SD | *N* | Mean | SD | *N* | Mean | *N* |
| XII | 39.1 | 9.8 | 32 | 33.3 | 11.7 | 64 | 22.6 | 16.9 | 25 | 31.2 | 121 |
| XI | 35.1 | 11.9 | 47 | 35.9 | 7.6 | 29 | 23.0 | 16.1 | 26 | 29.6 | 102 |
| X | 36.6 | 6.8 | 45 | 31.4 | 10.2 | 43 | 28.7 | 9.6 | 59 | 30.5 | 147 |
| IX | 38.6 | 8.2 | 30 | 26.5 | 8.7 | 27 | 21.9 | 10.3 | 33 | 28.5 | 90 |
| VIII | 31.9 | 8.3 | 57 | 29.4 | 7.5 | 55 | 22.0 | 15.4 | 29 | 27.9 | 141 |
| VII | 27.3 | 7.1 | 68 | .... | .... | ... | 20.5 | 7.0 | 63 | 25.6 | 131 |
| VII–XII | 33.0 | 9.8 | 279 | 32.4 | 9.6 | 218 | 23.4 | 12.2 | 235 | 30.0 | 732 |

## COMPARISON OF GRADES IN FACTUAL MEMORY

*Grade Averages.*—The data presented in the previous Tables (I–V) enable us to make the following comparisons of the grades participating in these experiments.

In immediate memory the averages increase quite markedly from grade to grade. Of the fifteen comparisons that may be made on the basis of the results of the various materials eight of the differences turn out to be four times their probable errors or more and are, therefore, completely reliable. Three more comparisons are differences 1.8 times their probable errors, or better, the chances of a true difference greater than zero being at least ninety in one hundred. A comparison of the A-Score means reveals an average increase of 3.05 per cent from grade to grade, all the differences being reliable.

In memory delayed one day the grade averages again increase as in Experiment 1. Of the fifteen comparisons six give reliable differences, and four more are more than 2.5 times their probable errors. The A-Score differences between the means of the grades average 4.5 per cent, but all are not completely reliable.

In the experiments with intervals of fourteen days and longer, no relationship is to be noted between factual memory and grade, the rank order of the grades being haphazard and most of the differences

being negligible. Whereas the means of the seventh and twelfth grades in the immediate memory and one day delayed memory experiments are well set apart by a distance of about 1.5 standard deviation of the entire distribution, the differences between the lowest and highest averages in the experiments with longer intervals are small. It may be inferred that groups originally quite different with respect to knowledge of an article read a single time become more and more alike with lapse of time—*i.e.*, the factors influencing forgetting operate differently on groups of differing abilities.

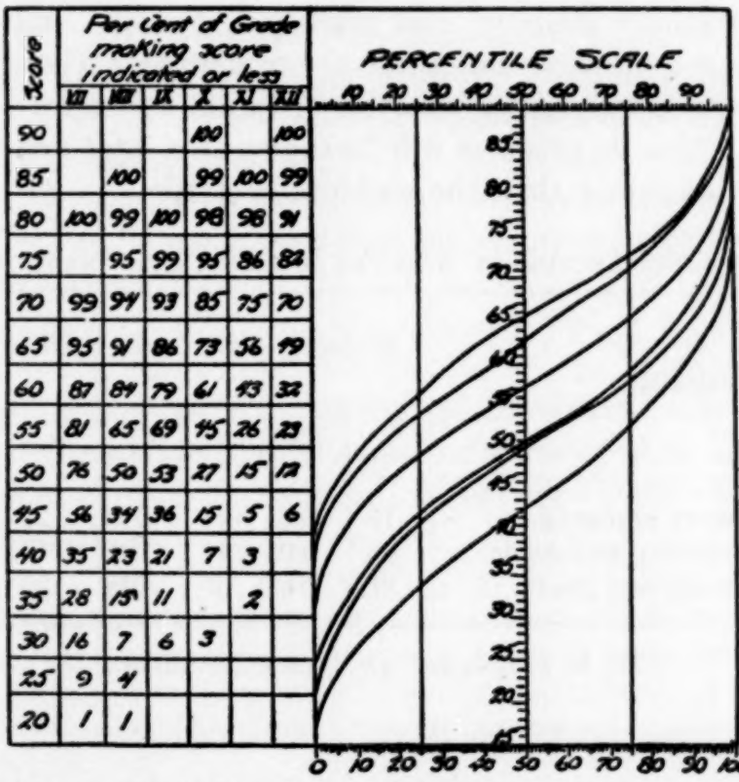



FIG. 2.—Percentile curves showing the overlapping of grades in immediate memory for Arkwright.

*Overlapping of Grades.*—While the average score in immediate and one day delayed recall increases perceptibly with grade, nevertheless adjacent grades overlap to a considerable extent. We find, *e.g.*, that the seventh, eighth, ninth, tenth, eleventh and twelfth grades reach or exceed the mean in immediate memory of Radium for all grades taken together by the following amounts respectively: 33.6 per cent, 25.0 per cent, 40.3 per cent, 57.0 per cent, 58.6 per cent, and 86.8 per cent. Again, 85.6 per cent of the distribution of the combined grades for Radium in immediate memory fall within a range common

to all six grades, and, similarly, 87.2 per cent of the combined scores in Germans and 82.6 per cent of those in Arkwright. The degree of overlapping is strikingly brought out by the percentile graph for immediate memory of Arkwright, Fig. 2, from which may be read the percentage of scores of any grade which fall below or above any given point in the distribution of any other grade. The curves have been smoothed free-hand.

*Correlations.*—The product moment correlations obtained for the factual memory results of several of the groups with grade in school are presented in Table VI. It will be seen that the correlations between immediate memory and grade are positive and low, those between delayed memory and grade are negligible. These results are in accord with the foregoing analysis of grade averages. The slight correlations which do exist, as will be shown in a later article, are due to other factors rather than the accident of grade.

TABLE VI.—CORRELATIONS OF FACTUAL MEMORY AND GRADE IN SCHOOL

| Variables | Radium | | Germans | | Arkwright | | Average |
|---|---|---|---|---|---|---|---|
| | *r* | PE | *r* | PE | *r* | PE | *r* |
| Immediate memory and grade..... | .49 | .035 | .24 | .041 | .29 | .045 | .34 |
| Fourteen day memory and grade.... | .04 | .049 | .34 | .040 | −.07 | .043 | .10 |
| Thirty day memory and grade..... | .11 | .043 | .07 | .049 | .36 | .090 | .18 |

(*To be concluded in December issue*)

# WHY OTIS' "IQ" CANNOT BE EQUIVALENT TO THE STANFORD-BINET IQ

PSYCHE CATTELL

Harvard University

Dr. Otis has deviced a measure of brightness for use with his Self-administering Tests which is derived by determining the number of points in score the child varies from his age norm, if he is above the norm this amount is added to one hundred, if below subtracted from one hundred. Dr. Otis states that when an IQ has been found by other means than the Binet Tests the term "IQ" must not be used without qualification, yet he writes that "it seemed best simply to call this new measure of brightness an 'IQ.'" He apparently justifies the use of the term "IQ" on the erroneous belief that his measure of brightness is equivalent to the Binet IQ and that the group test IQ's obtained by the standard method are not.

Dr. Terman found the middle fifty per cent of the Stanford-Binet IQ's to have a range from ninety-two to one hundred eight. Dr. Otis found the middle fifty per cent of the scores made in his tests to have approximately the same range in the several age groups, that is that fifty per cent of the cases fell within eight points of the norm, he then concluded:

> Fortunately, therefore, each point in the score of an individual above or below the norm for his age represents a point in IQ above or below one hundred. If an individual score exceeds the norm for his age by twelve points, his IQ is one hundred twelve . . . This being the case it seemed best simply to call this new measure of brightness an "IQ."

Dr. Otis has failed to take into account the fact that two distributions may have the same medians and the same middle fifty per cent range and still differ widely at the extremes. It can be shown both theoretically and empirically that there is a constant difference between the Binet IQ and Otis' "IQ" at the extremes, especially the upper extreme. That this must of necessity be true can be demonstrated from the data given in the manual of directions for the "Otis Self-administering Tests of Mental Ability." An inspection of the table of norms to determine the highest and lowest "IQ" that it is possible for a child of a given age to obtain by this method will show the impossi-

bility of the range of Dr. Otis' measure being equal to that of the Binet IQ.

While Dr. Otis makes no mention of the relative reliability of his table of norms at different ages, it is perhaps, not fair to expect too great accuracy at the extremes. Therefore, the ages of 10-6, 13-0 and 15-6 have been selected for a comparison of the Binet IQ and the "IQ" obtained by Dr. Otis' method. These ages are the quartile points of the age range of the table of norms and should include the more reliable part. The curve to the left of the accompanying chart is the frequency distribution of the IQ's of Professor Terman's original one thousand unselected Stanford-Binet IQ's. The curve to the right is a frequency distribution of the Binet IQ's of his six hundred forty-three gifted children, largely from Grades III to VIII and all below the age of fourteen years. The lowest and highest Otis "IQ's" that it is possible for children of a given age to obtain on the Intermediate Self-administering Test are indicated on the IQ scale at the foot of the chart.

The highest possible Otis "IQ" that could be obtained by a child aged 13-0 on this test is one hundred thirty-one and the lowest fifty-six. This is a narrower range than that of the Stanford-Binet IQ's. If Professor Terman had used this test and method in selecting his gifted group he might have found one or two children that measured up to one hundred thirty but none above one hundred thirty-one (his group of six hundred forty-three contained only one case with an IQ below one hundred thirty-five). At the age of 10-6 the highest possible Otis "IQ" is one hundred forty-eight (approximately one half of Professor Terman's gifted group had IQ's of one hundred fifty or higher). The lowest a child could attain at this age is seventy-three. Every large school system has a number of children with Binet IQ's below seventy, At age 15-6 the highest "IQ" obtainable is only one hundred twenty. approximately two standard deviations above the norm. Even if the advanced test were substituted for the intermediate the highest possible "IQ" would still be only one hundred thirty-seven. The lowest "IQ" obtainable at this age is forty-five, even this is not as low as the Standford-Binet IQ's frequently met with. At no age is there a possible Otis "IQ" range of more than seventy-five points. At the age of 12-0 it is theoretically possible to get a range from sixty-three to one hundred thirty-eight. A younger child could get a higher Otis "IQ," but could not make one as low as sixty-three, while an older child could make one lower than sixty-three,

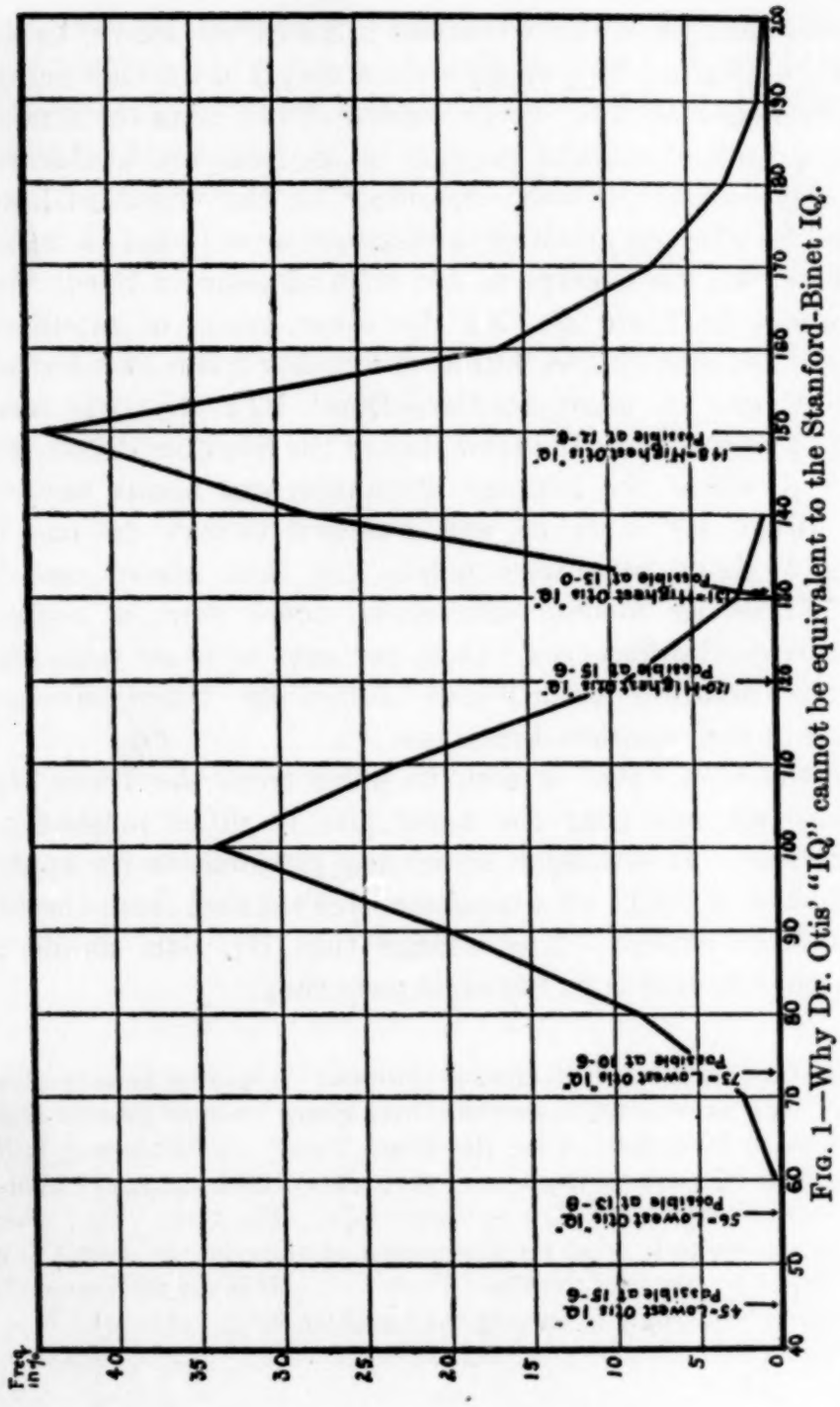



FIG. 1—Why Dr. Otis' "IQ" cannot be equivalent to the Stanford-Binet IQ.

but would have no possibility of making one as high as one hundred thirty-eight.

Similar conclusions were reached in an earlier study[1] based on the actual IQ's obtained by giving Forms A and B of the Otis Self-administering tests and at least two Standard-Binet tests to three hundred twenty-two school children ranging in age from ten to thirteen inclusive. Those children who according to the Stanford-Binet tests possessed an average grade of intelligence were found to be about six points lower on the average in Dr. Otis measure in Form A and two points above in Form B. At the lower levels of intelligence the differences are slightly less, but at the higher levels as selected by the average of two or more Stanford-Binet IQ's the Otis measure is markedly and consistently below that of the Stanford-Binet, especially in Form A where the average difference was about twelve points at the Binet IQ level of one hundred twenty to one hundred thirty and about seventeen points for that above one hundred thirty. These are median differences, there were, of course, many larger individual differences, fifteen per cent of those with Binet IQ's above one hundred twenty-four varied by twenty-five or more points, all in the negative directions.

Thus the Otis "IQ" is seen to differ from the Binet IQ by an appreciable amount near the norm and to differ markedly at the upper extreme. It is difficult to see any justification for applying the unqualified term "IQ" to a measure which differs from the Standard IQ to such an extent. It is strange that Dr. Otis should propose doing so when he writes on the same page that:

> It seems that the term "Intelligence Quotient" is coming to have a legal recognition, but IQ's as sometimes derived from group tests of mental ability bear little relation to IQ's derived by the Binet Tests[2] . . . Unless it is distinctly understood how IQ's were derived . . . they should be designated by some means ~~such~~ as National IQ's, Otis IQ's or Binet IQ's. The term "IQ" when not so qualified or understood, must be interpreted as referring to actual Intelligence Quotients found by means of the Binet Tests.[3] . . . It is the purpose of the author to use the term "IQ" only in its original significance.[2]

---

[1] CATTELL, PSYCHE: IQ's and the Otis Measure of Brightness. *Journal of Educational Research*, June, 1930.

[2] "Otis Self-administering Tests of Mental Ability." Revised Manual, p. 5.

[3] "Statistical Methods in Educational Measurements." World Book Co., Yonkers, New York, p. 156.

Yet in proposing his new measure of brightness which has just been shown to differ from the IQ both in its meaning and method of derivation he apparently believes that he is adhering to his statement that "It is the purpose of the author to use the term 'IQ' only in its original significance" when he writes: ". . . It seemed best simply to call this new measure of brightness an 'IQ.'"

# PROGNOSIS OF ABILITIES TO SOLVE EXERCISES IN GEOMETRY

WINONA M. PERRY

University of Nebraska

What are the abilities essential to the solving of the various exercises confronting the student of plane geometry? Since it is possible that these abilities may be component parts of "general intelligence," analysis of intelligence tests might be the means of detecting the constituent abilities. Accordingly, twenty-six group intelligence tests (or forms) were examined and their content observed as verbal, numerical, or spatial in character. The separate elements, and subtests too, were analyzed further into (1) abilities to perceive singly (as in recognition of the name, size or location of an object), or to perceive several objects for the purpose of comparisons of similar or non-similar meanings, shapes, or amounts (as in scaling to denote relative amounts of beauty); into (2) abilities to recall word meanings, attributes, and relationships; and into (3) abilities in purposive thinking: To discover relationship (as in number series, analogies, and mixed relations, disarranged elements, completion of sentences and pictures), to judge the possibility and correctness of statements, to apply a principle singly, and to weight elements (as in finding the best answer, the correct pathway, and in dealing with certain elements inferentially).

Which of these abilities are selected and organized into the series of changing responses occurring during the processes of solving geometric exercises? Perception is essential; perception, in terms of the given figures as a whole or in parts, is dependent upon the recall of meanings of figures, of relationships, and of properties; *i.e.*, a sufficient degree of meaning or familiarity with vocabulary units is necessary to note and to follow commands or questions in terms of the total figure as given, or its parts, or the relationship between these parts. Not essential, but usually following this series of perceptions is their motor expression in the drawing of the figure described and in the stating of the facts given and those which are to be proved; this combination of responses results in perception of known relationships in the figure as drawn. Recall in the light of analysis means pertinent recall which immediately makes available the needed meaning of certain figures previously reacted to, or their nature, relationships, or properties. Very complicated is the series of responses, "purposive thinking,"

which is to determine the link that connects the conclusion—to be judged, drawn, or proved—with the recalled and analyzed facts. The more major responses included under purposive thinking may be indicated by these questions:

Which elements among those recalled "fit in" with the conclusion?

Which elements among those analyzed in the light of the figure "fit in" with the conclusion?

Which ones among the pertinent recalled elements differ from, or are not included among the pertinent analyzed facts?

Is evidence available to help in finding *the one* element necessary to complete the identity between the series of analyzed and of recalled facts?

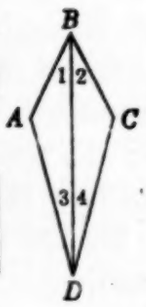
$S \longrightarrow R_1$

[If a diagonal of a quadrilateral bisects those angles whose vertices it joins, the diagonal divides the figure into two equal triangles.] [Meaning of statement as a whole]

$R_{11}$ [Perception of conclusion, that triangles are equal.]
$R_{12}$ [Perception of total figure, a quadrilateral.]
$R_{13}$ [Perception of relationship between parts of figure, diagonal bisecting opposite angles.]

$R_{11}$, $R_{12}$, $R_{13}$ ——→ $R_2$ Motor Expression

$R_{21}$ [Drawing the figure.]
$R_{22}$ [Stating what is given. To be proved.]




→
$R_3$ [Perception of known relationships in figure.]
In △ *ABD* and *CBD*,
$R_{31}$ ∠1 = ∠2 Given.
$R_{32}$ ∠3 = ∠4 Given.
$R_4$ [Recall in the light of analysis.]
Conditions for equal triangles.
$R_{41}(s - a - s)$
$R_{42}(a - s - a)$
$R_{43}(s - s - s)$

$R_3\begin{pmatrix}R_{31}\\R_{32}\end{pmatrix}$, $R_4\begin{pmatrix}R_{41}\\R_{42}\\R_{43}\end{pmatrix}$ →$R_5$ [Which elements of $R_3$ and $R_4$ "fit in" with $R_{11}$? Two pairs of equal angles, in $R_3$ and $R_{42}$.] →$R_6$ [Which element, or elements in $R_{42}$ differ from those in $R_3$? Included side.] →$R_7$

$R_7$ [Does *BD* in △*ABD* equal *BD* in △*CBD*? Yes, "common."] →$R_8$ [Conclusion of equal triangles proved. $(a - s - a)$ Thus, the common side, *BD*, is the unifying or link factor between the desired, and given and recalled facts.]

Does this completion of the identity between the two series of facts result in a unified grouping, or classification, or "gestalt" of all

the facts concerned? Furthermore, does recognition of success follow the solution of the exercise?

These analyses are illustrated in the exercise shown on p. 605.

These analyses led to the construction of a test prognostic of abilities in plane geometry. The situations and the responses to those situations are numerical and verbal, as well as spatial in content. The number and grouping of elements are to test the complex, organized abilities and certain of their constituent parts, in order to reveal the status of each student's abilities in perception, in recall, and in purposive thinking; they serve an additional purpose, also, that of differentiating between errors due to carelessness and to inabilities or low degrees of these abilities. The students' strengths and weaknesses thus detected may become the guides toward differentiated teaching emphases. The test includes elements in the form of directions to be followed, detecting of similar figures, recall, drawing and judging of conclusions, detection of correct reasons from classified series of statements for stated steps in geometric exercises, and motor expression of spatial imagination.

While this test was being constructed, another test appeared—"Orleans Geometry Prognosis Test." Although the purpose of each is identical, the philosophy underlying the construction of each and the nature of the content are strongly contrasting. The philosophy of its authors (Orleans) seems to be that prognosis of student abilities to learn geometry is best secured by student responses to varied samplings of geometric situations; *i.e.*, that prognosis of how students will learn geometry is by means of how they do learn when responding to actual geometric situations. On the contrary, the philosophy underlying the test (Perry) first described is that the more efficient prognosis will be effected by means of tapping the students' potential abilities which, when directed, can lead to effective mastery of plane geometry. The "Orleans Geometry Prognosis Test" requires: Detection of identical meanings, of named angles, of angle size, of complementary and supplementary angles, of relative angle and side positions, of equal angles and of equal sides and double lengths of line-segments; perception of statement meanings due to perception of marked equal parts; mere detection of "if" and "then" clauses; and purposive thinking due to the selection of correct reasons from a given series of unclassified statements.

Both tests were given, though in reverse order, to students in two classes in plane geometry in Lincoln High School during the second

semester of 1929–1930. Since the papers of students who were repeating the course or who were absentees on one day could not be included, there were forty-two students whose responses to the elements of each test were studied in detail and comparatively. In addition, there were two methods used for measuring the achievement of these students following their study of Book I: One was the study of student responses to Test 1 of "Hart's Geometry Tests," a test which is a part of the regular class work in this school; the other means were the grades assigned by the class instructors, although the teachers' estimates of student achievement were influenced to some extent by the responses of these students to the Hart test. Students' relative positions according to their total scores on each prognostic test, also their IQ's,* were charted with their achievement positions indicated by total scores on the Hart test and by instructors' grades. Two techniques, that of the Pearson Product Moment Method and that of the Pearson Coefficient of Mean Square Contingency, were used to measure the relative amounts of correlation detected by each of these three methods of prognosis and two methods of determining achievement. The total *r*'s are stated in the first part of Table I; the partial *r*'s, with IQ's "partialled out," in the second part.

TABLE I

| | Total *r*'s between methods of determining prognosis and achievement in geometry | | | Partial *r*'s, (*r*'s from which IQ's have been partialled out) | |
|---|---|---|---|---|---|
| | Perry | Orleans | IQ | Perry | Orleans |
| Grade..... | .61 | .66 | .48 | .49 | .50 |
| Hart..... | .63 | .78 | .35 | .56 | .42 |
| IQ..... | .50 | .76 | | | |

From these tables, and especially from the latter part of Table I, it appears that Orleans' test draws the more heavily upon intelligence as estimated by the Terman Group Test of Mental Ability. This is noted from the drop of .12 in the measurement of the correlation, when IQ's have been partialled out, between the Perry test and the teachers'

* The IQ's, obtained from the files in the office of the high school, were determined by the Terman Group Test of Mental Ability.

grades, and similarly a drop of .16 between Orleans' test and teachers' grades. When achievement is determined by the scores on the Hart test, there is a loss of .07 in the partial "*r*" from the total "*r*" between scores on the Perry and on the Hart tests, but similarly a much larger loss of .36 between scores on the Orleans and Hart tests. Furthermore, since the partial *r*'s (.49 and .50) are nearly identical, but the partial *r* (.56) is much larger than the partial *r* (.42), the Perry and the Orleans tests seem to prophesy achievement equally well *if* achievement is estimated by teachers' grades, but prognosis from scores on the Perry test is the more efficient *if* achievement is measured by scores on the Hart test. (Student time required for responding to the Perry and to the Orleans tests was forty-five, and seventy minutes, respectively.)

As the number of cases is small, further analysis of correlation, as detected by three methods of prognosis and two methods of achievement, was accomplished by means of the Pearson Coefficient of Mean Square Contingency. Approximately equal proportions of the range of scores on the Perry and the Orleans tests, and those parts of the range of IQ's as used in the differentiation of class members, were compared with the teachers' grades,* as interpreted by the school administrators, and with the published quartiles of scores on the Hart test; these appear in Table II.

TABLE II.—STUDENT PLACEMENT ACCORDING TO PROGNOSIS AND ACHIEVEMENT

| Test | Scores or IQ's | Teacher's grades | | | | Hart scores | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 7 | 6–5 | 4–3 | 2–1 | 27–33 | 34–40 | 41–45 | 46–50 |
| Perry | 62–77 | .. | .. | 11 | 10 | .. | 1 | 8 | 12 |
| | 50–61 | 2 | 5 | 6 | 1 | 1 | 4 | 6 | 3 |
| | 42–28 | 1 | 2 | 4 | .. | 1 | 4 | 1 | 1 |
| Orleans | 80–140 | 1 | 2 | 14 | 9 | .. | 2 | 10 | 14 |
| | 40–79 | 1 | 3 | 6 | 2 | 1 | 4 | 5 | 2 |
| | 8–39 | 1 | 2 | 1 | .. | 1 | 3 | .. | .. |
| IQ'S | 120–140 | .. | .. | 3 | 5 | .. | .. | 2 | 6 |
| | 95–119 | 2 | 3 | 18 | 6 | .. | 6 | 13 | 10 |
| | 79–94 | 1 | 4 | .. | .. | 2 | 3 | .. | .. |

* The grades, (1 and 2, 3 and 4, 5 and 6, and 7) are interpreted in per cents (in the ninties, eighties, seventies and below seventy, or "failure"), respectively.

The Pearson Coefficients of Mean Square Contingency,* computed from Table II, are stated in Table II*a*. (The amounts of these coefficients could not exceed .866.*)

TABLE II*a*.—$C_2$'s BETWEEN PROGNOSIS AND ACHIEVEMENT

| | Perry | Orleans | IQ |
|---|---|---|---|
| Grade | .55 | .43 | .61 |
| Hart | .51 | .57 | .63 |

The differences between the $C_2$'s of the Perry scores and the measures of achievement, and the corresponding $C_2$'s of the Orleans scores and of the IQ's, range from −.12 through .06 to .12. Thus, although there is association between each method of prognosis and of achievement, no marked superiority in their measurement was shown by any one of the techniques used.

In conclusion: Prognosis of abilities to solve exercises in geometry seems the more efficient when based upon analysis of the requisite abilities and their constituent parts. This means of prognosis (Perry test) not only requires less student time but it also duplicates less those abilities interpreted by the IQ's, than does the Orleans test. Scores on the Perry Geometry Prognosis Test are proving indicative of student status in achievement, and furthermore of information concerning student strengths and weaknesses—information of paramount importance to the teacher who is to direct the learning of the students of plane geometry.

* The technique for computing these coefficients is described in "An Introduction to the Theory of Statistics" by G. Udny Yule (pp. 66–67), also in "Statistics in Psychology and Education" by Henry E. Garrett (pp. 195–203).

# THE DEVELOPMENT OF MENTAL ABILITY AT THE COLLEGE-ADULT LEVEL

MELVIN B. WRIGHT

University of Pennsylvania

If we may define behavior as the response of an organism to a stimulus situation, it is possible that a more complete knowledge of the organism, may enable us to make more accurate predictions, with respect to subsequent behavior. Experience has shown that there is a high degree of correlation, between the status of mental development and the quality of the personality products.

Most of the investigations of Mental Ability, have been confined to ages ranging from three to eighteen years. This is partially due to lack of materials and methods for investigating the mental growth at the higher age levels. Furthermore, while it has been generally assumed that biological and environmental factors both contribute to mental development, but little progress has been made in analyzing the contributions of either.

The present investigation was undertaken, in order to determine what progress was taking place in mental development, at the College-Adult Level, an age range from eighteen to twenty-three years. Secondly, we propose to analyze this mental growth in such a way as to reveal the relationship between its biological aspects, and those which may be attributed, more directly to environmental pressures.

Terman has pointed out that though the Binet scale does not properly evaluate children who test at one hundred thirty, beyond the fifteen year age limit; this "does not mean that development ceases at this time, but merely that the Stanford-Binet does not measure it."[1]

Thorndike gives a parabolic curve, resembling the learning curve, rising rather abruptly from six to fifteen or sixteen years and then approaching a plateau which he describes as, "the probable form of the curve of intellect in relation to age."[2]

Thurstone has shown a mental growth curve, based on Binet test data, for ages three to fourteen years, concerning which he says, "A striking feature of this curve is that it continues to rise even at the age

---

[1] Terman, L. M.: "The Intelligence of School Children." P. 147.

[2] Thorndike, E. L.: "The Measurement of Intelligence." P. 5.

of fourteen years, with no indication of reaching a level. It certainly looks as tho the kind of intelligence which is measured by Binet tests and their variations, continues to grow as rapidly at fourteen years as it does at nine. This conclusion contradicts the statement frequently made, to the effect that test intelligence approaches an adult and more or less stable level at fourteen to sixteen years. The appearance of these curves, indicates that the growth of test intelligence continues beyond the age of fourteen years . . . Common sense judgment certainly favors the assumption that the average man of forty years is more intelligent than the average boy of twenty, but so far we have not been able to measure that difference. Instead of acknowledging this limitation in our measurement methods, we have not infrequently attempted to juggle with the definition of intelligence, to make it fit the measuring devices that are accessible."[1]

There are three curves in the graph to which Thurstone refers, all approximately parallel, one of which represents the mean ability of successive age groups from three to fourteen years on Binet tests. The second represents those who had a plus one sigma rating above the means of their respective age groups, and the third represents those who had a minus one sigma rating, below the mean for their respective age groups. All these curves show a correlation of better than .97 between test ability and age.

In a later article, Thurstone has extended this curve, by using similar data, for ages three to seventeen years, and interpolating the curve back to birth. The main part or central extension of the original curve, from three to fourteen years is slightly modified, so that in the later curve, an inflection points develops between the ninth and twelfth years, possibly about the eleventh year. From birth to the third year, the curve shows positive acceleration, which continues to the eleventh year, while from the fourteenth to the eighteenth year, the acceleration becomes about equally negative. This curve is shown in the diagram attached to this paper.

Thurstone concludes that the mental growth curve is asymptotic both to absolute zero and to an adult level. The lower limit is established at −.43 at birth; the upper limit is not ascertainable by computation. "These findings contradict the hypothetical mental growth

---

[1] Thurstone, L. L.: A Method of Scaling Psychological Tests. *Journal of Educational Psychology*, Vol. XVI, No. 7, p. 442.

curve of the text books, which is usually shown to have negative acceleration from birth without an inflection point."[1]

The present investigation attempts to measure the development of mental ability that lies beyond the range of Binet scales, as reported by Thurstone. It further seeks to ascertain whether mental development has reached the plateau stage between the ages of seventeen and twenty-three, at the College-Adult Level.

Thurstone regards his second mental growth curve, as giving a "better picture of continuity for mental growth regarded as a biological function." In pursuit of similar information, we have chosen a battery of tests, which we believe will give an index of the continuance or decline of this biological phase of mental growth. We have characterized this phase as Fundamental Ability.

We have also chosen a second battery of tests, which will evaluate another phase of mental growth, which may be defined as Complex Mental Processes. The former may be construed as concomitant with biological growth, while the latter represents the effect of educational and cultural pressures. The same tests were presented to all ages from eighteen to twenty-three on the assumption that there was equal competency within the group. This assumption was justified, in that all ages scored practically over the same ranges, in all tests, with varying degrees of frequency.

The results were obtained in the psychological laboratory at the University of Pennsylvania, from college men and women, during the three academic years, 1927–1928–1929–1930, as part of the required work in psychology courses. This testing program was instituted by Dr. Lightner Witmer, and has been developed over a number of years, by instructors in the clinical laboratory sections of first year psychology courses.

The tests, to which the students were subjected, were used as didactic experiments or demonstrations. They were administered under standard conditions and procedure, during the course of twelve class sessions, by a number of instructors, and scored on established standards. The significant measures for each test were derived from a combined protocol, of the various groups that had taken the tests. More than eleven hundred students have been tested in this manner during the three years, and have contributed to these results.

---

[1] Thurstone, L. L.: The Mental Growth Curve for Binet Tests. *Journal of Educational Psychology*, Vol. XX, No. 8, p. 569.

Dr. Brotemarkle, under whose supervision these courses were given, conceived the idea of using the scores to develop a "Mental Graph," by means of which various types of student ability might be portrayed graphically.[1] One of the distinctive features of this Mental Graph, is the analysis of Mental Ability into two parts, which he designates as: (1) Fundamental Abilities; and (2) Complex Mental Processes. In evaluating any personality, or its products, it is manifestly of great importance, from both the diagnostic and prognostic points of view, to know whether an individual's inferiority or superiority, can be traced to either one or the other of these two aspects of his Mental Ability. Measurement on the Stanford-Binet, and other group or individual tests, affords no such insight.

The pragmatic vindication of this procedure is evidenced by the types of graph actually obtained in practice, which Dr. Brotemarkle divides into the following seven classes:

1. Superior—meaning superior in Fundamental Ability and in Complex Mental Processes.
2. Superior descending—meaning superior in Fundamental Ability descending in Complex Mental Processes.
3. Median ascending—meaning median in Fundamental Ability and ascending in Complex Mental Processes.
4. Median—meaning median in Fundamental Ability and Complex Mental Processes.
5. Median descending—meaning median in Fundamental Ability and descending in Complex Mental Processes.
6. Inferior ascending—meaning inferior in Fundamental Ability and ascending in Complex Mental Processes.
7. Inferior—meaning inferior in both Fundamental Ability and Complex Mental Processes.

It is by no means inevitably assured that an individual with a superior endowment of Fundamental Ability, will perforce remain superior in the field of Complex Mental Processes. Nor is it assured that a median endowment of Fundamental Ability, may not rise or fall in the field of Complex Mental Processes. From these findings, it became apparent that any estimate of the growth of mental ability at the College-Adult Level, must be analyzed, not only with respect

---

[1] Brotemarkle, R. A.: College Student Personnel Problems. *Journal of Applied Psychology*, Vol. XI, No. 6.

to increasing age, but also with respect to the development of these two phases of mental ability.

Our investigation concerns the achievements of the various age groups from seventeen to twenty-three, comprising the College-Adult Level, in these two fields of Mental Ability. In forming the age groups, each one was made to include all individuals who had passed a given birthday, but had not passed the age level of the next higher group. Thus the eighteen year old group comprises those who have passed their eighteenth birthday, but have not yet reached the nineteenth.

The results obtained by the seventeen year old group, are not to be regarded as highly significant, owing to the very few cases available at this age level. The twenty-three plus group, represents a small miscellaneous company, ranging from twenty-three years to as high as 40 years. Their heterogeneity and limited number place a high degree of unreliability on the results obtained by them. These facts must be constantly borne in mind, when reading the extremities of our subsequent tabulations.

More than eleven hundred students were tested, though not all of these took all of the tests, due to absences and interruptions of the laboratory sessions. From this mass of material, we had available, between six hundred and seven hundred scores for each test, about equally divided between men and women. The scores were grouped in respective protocols for each age level. A subsequent investigation of these results will be conducted, on the basis of sex, to ascertain what differences may prevail between them.

Wherever we had a score, belonging to an individual of a particular age group, it was recorded, regardless of how many of the twelve tests, the individual may have missed. Our final results are therefore more significant, from the point of view of achievement in relation to age, since the results of no two tests, include scores by exactly and exclusively the same individuals. Nearly all the students took almost all of the tests, but some who took test number two, for example, may have missed tests three and five and *vice versa*. The results therefore, are based on scores, belonging to certain age groups, rather than to certain groups of persons of a given age. This factor dispels, to some extent, the possibility of a fortuitous selection of subjects. Ours is a random sample of the College-Adult Population.

Following Brotemarkle's analysis of Mental Ability into the component parts of Fundamental Ability and Complex Mental Processes,

the six tests which were used to measure each one of these factors, in this investigation, are listed below:

| Fundamental Abilities, Names of Tests | Complex Mental Processes, Names of Tests |
|---|---|
| 1. Memory Span for Digits | 1. Opposites (Roback) |
| 2. Memory Span for Syllables | 2. Abstraction (Roback) |
| 3. Memory Span for Ideas | 3. Reference (Roback) |
| 4. Taylor Number Test | 4. Definition (Brotemarkle) |
| 5. Aussage (Undirected Attention) | 5. Judgment (Roback) |
| 6. Ausfrage (Directed Attention) | 6. Trabue Language Scale J |

These tests have been presented individually by their authors and subjected to analysis by Brotemarkle[1] and Miller.[2] It is sufficient to indicate their general characteristics and the differences between the two batteries.

The battery used for evaluating Fundamental Abilities, seeks to ascertain the relative plasticity of the organized protoplasm, its impressionability. The memory span tests are almost as simple as the patellar reflex, but limits of achievement vary with the individual, as does the excursion of the foot in the knee jerk. The remaining three in this battery, are used to explore the field of attention. Is the individual alert and receptive to the world about him? To what extent does the world intrude itself upon his personality, and make impressions upon him. These three offer another clue to the individual's plasticity, his impressionability, the ease or difficulty with which he can be modified or adjusted to his environment.

The tests for Complex Mental Processes, investigate a different aspect of mental function in personality. Here we observe and seek to evaluate, the influence and results of training and environment, the height of the cultural level, the clearness or obscurity of thought processes, the power of abstract thought, the ability to construct and compare concepts, keenness of discrimination, the precision with which one may describe or define, and so on. The whole range of complex mental functions in personality invites our inquiry, but at best, we can only select a few samples in the directions indicated above.

---

[1] Brotemarkle, R. A.: College Student Personnel Problems. *Journal of Applied Psychology*, Vol. XI, No. 6.

[2] Miller, Karl G.: The Competency of Fifty College Students. *The Psychological Clinic*, Vol. XIV, Nos. 1–2.

We have treated the results obtained by these two batteries, separately. The scores for each test for each age group, were tabulated and the average and sigma determined. A composite score for each age group, for each battery, was then developed. These composite scores are known as the Fundamental Ability Index and the Complex Mental Processes Index.

The method of obtaining these composite scores consisted first of all in weighting each of the six tests of each battery separately. The amount by which each test should be weighted was determined by the ratio, which its sigma bore, to the average sigma of the six tests in the battery. That is, the average sigma of the six tests, was divided by each of its constituent sigmas, and the resulting quotient was the weight chosen for that test. The mean of each test was multiplied by its respective multiplier; the results summated, and an average mean for the six tests derived, forming a composite score or index.

We have designated this statistical procedure as the average sigma method of weighting or combining test scores. Each of the six constituent elements of the composite mean, bears the same mathematical relation to the original from which it was derived, that its respective sigma bears to the average sigma of the battery.

This method, offered a consistent scheme for treating many combinations of scores in both batteries, and further eliminated the possibility that obtained results might have been influenced by a fortuitous choice of weights for the various tests. These indices were then subjected to the statistical analysis indicated in the discussion of results. Correlations were obtained by the Pearson Product Moment Formula.

A method of combining test scores, somewhat similar to that used in this investigation, is described by Garrett.[1] The "Chance $R$" for multiple Correlation, was obtained by the following formula:[2]

$$\text{"Chance } R\text{"} = \frac{\sqrt{n-1}}{\sqrt{N}}$$

## Analysis of Results

The means attained by the various age groups in the tests which constitute the battery for Fundamental Abilities, are found in Table I.

[1] Garrett, H. E.: "Statistics in Psychology and Education." P. 280.
[2] Garrett, H. E.: "Statistics in Psychology and Education." P. 239.

In as much as the populations of the various age groups differed considerably, the greater proportion of mathematical reliability belongs to the group having the largest population.

The scores of the seventeen year age group, have been discarded in subsequent computations, because their small number gives an exceedingly high degree of unreliability to these scores. Furthermore, they represent a highly selected group of individuals, who by reason of an early start, special teaching, ability to skip a grade, or some other reason, reached the University before their eighteenth birthday. The size of this group as compared with the eighteen, nineteen, and twenty year group, indicates that some factor of selection was at work. If this factor happened to be superior mental ability, this marks them all the more as a highly selected group of seventeen year old individuals. For our purposes we need random samples of all ages investigated, in sufficient numbers to give some degree of reliability.

The group listed as twenty-three plus, represents miscellaneous ages as high as forty and is therefore not indicative of any age. The scores of the twenty-two year old group have been used, though they number but twice the size of the seventeen year old group. However, they do not indicate that special factors contributed to their selection, which were not also applicable to ages eighteen to twenty-one. Nevertheless, due consideration must be given their small population in evaluating their indices.

TABLE I.—SHOWING THE AVERAGE AND STANDARD DEVIATION FOR EACH AGE GROUP IN FUNDAMENTAL ABILITIES

| Age | Frequency | MS digits | | MS syllables | | MS ideas | | Undirected attention | | Directed attention | | Taylor number test | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Average | SD | Average | SD | Average | SD | Average | SD | Average | SD | Average | SD |
| 17 | 19 | 9.79 | 1.78 | 32.05 | 3.34 | 5.20 | 1.77 | 10.21 | 3.94 | 24.66 | 4.77 | 14.52 | 3.36 |
| 18 | 109 | 8.62 | 1.57 | 33.48 | 4.38 | 5.28 | 2.14 | 9.85 | 3.30 | 22.28 | 5.13 | 14.17 | 4.24 |
| 19 | 236 | 8.48 | 1.58 | 29.97 | 6.27 | 4.99 | 2.15 | 9.36 | 3.40 | 23.57 | 5.16 | 13.66 | 3.56 |
| 20 | 196 | 8.22 | 1.55 | 30.22 | 5.43 | 4.68 | 1.99 | 9.44 | 3.74 | 23.24 | 5.55 | 13.31 | 3.62 |
| 21 | 100 | 9.35 | 1.55 | 29.79 | 6.03 | 4.77 | 2.03 | 10.12 | 4.32 | 22.92 | 4.95 | 14.22 | 3.74 |
| 22 | 38 | 8.89 | 1.56 | 30.93 | 6.15 | 4.97 | 2.06 | 9.31 | 4.50 | 22.20 | 6.48 | 14.32 | 3.82 |
| 23+ | 42 | 8.57 | 1.45 | 30.13 | 4.48 | 5.26 | 1.97 | 9.61 | 3.22 | 21.20 | 6.99 | 14.02 | 4.16 |

Table I gives little indication of the relative abilities of the respective age groups, by observation.

TABLE II.—SHOWING THE AVERAGE AND STANDARD DEVIATION FOR EACH AGE GROUP IN COMPLEX MENTAL PROCESSES

| Age | Frequency | Opposites | | Abstraction | | Reference | | Judgment | | Trabue | | Definition | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Aver-age | SD | Aver-age | SD | Aver-age | SD | Aver-age | SD | Aver-age | SD | Aver-age | SD |
| 17 | 18 | 49.88 | 10.65 | 51.05 | 12.65 | 30.52 | 4.92 | 60.20 | 12.90 | 11.00 | 1.67 | 47.36 | 12.90 |
| 18 | 98 | 49.67 | 12.85 | 51.74 | 14.35 | 28.68 | 7.02 | 57.40 | 14.50 | 9.52 | 1.78 | 48.53 | 15.05 |
| 19 | 226 | 50.00 | 14.15 | 47.50 | 15.70 | 29.65 | 5.97 | 55.83 | 12.75 | 9.89 | 1.61 | 43.32 | 15.65 |
| 20 | 185 | 48.59 | 13.35 | 51.96 | 15.35 | 28.50 | 5.43 | 57.42 | 13.40 | 9.49 | 1.54 | 46.10 | 14.85 |
| 21 | 93 | 50.73 | 13.35 | 55.06 | 15.30 | 29.09 | 5.82 | 55.00 | 13.40 | 9.24 | 1.58 | 45.79 | 19.35 |
| 22 | 33 | 46.66 | 15.20 | 54.22 | 21.40 | 29.62 | 6.64 | 59.08 | 12.65 | 9.60 | 1.67 | 46.34 | 12.85 |
| 23+ | 42 | 49.71 | 11.80 | 51.29 | 13.35 | 26.96 | 6.30 | 52.90 | 13.15 | 9.52 | 1.46 | 45.33 | 14.45 |

Table II shows in similar manner, the means achieved by the various age groups in the tests comprising the battery for Complex Mental Processes. The results do not greatly facilitate the discovery of trends in mean ability or variability, by observation.

TABLE III.—SHOWING THE COMPOSITE AVERAGE SCORE FOR EACH AGE GROUP IN FUNDAMENTAL ABILITIES

| Age | Frequency | Mean |
|---|---|---|
| 18 | 109 | 15.14 |
| 19 | 236 | 14.37 |
| 20 | 196 | 14.29 |
| 21 | 100 | 15.14 |
| 22 | 38 | 15.25 |

Table III shows the results of combining the six Fundamental Ability test scores, into a composite or index, for each group. The eighteen and nineteen year old groups appear to be superior to the twenty year old group, and the twenty-one and twenty-two year old groups are also superior to the twenty year old group. The average age of this entire College-Adult Level, weighted according to the number of each population, is 19.59 years, while the average mean for

the entire group is 14.63 in Fundamental Ability. The mean age of this group of College adults therefore lies between nineteen and twenty years, while the mean ability of the college adult is represented by a score that is closer to the score of these two ages, than it is to that of any other age group.

TABLE IV.—SHOWING THE COMPOSITE SCORE FOR EACH GROUP IN COMPLEX MENTAL PROCESSES

| Age | Frequency | Mean |
|---|---|---|
| 18 | 98 | 43.84 |
| 19 | 226 | 45.38 |
| 20 | 185 | 46.35 |
| 21 | 93 | 47.20 |
| 22 | 33 | 47.09 |

The results of combined scores for the Complex Mental Processes are shown in Table IV. Here the trend toward increasing ability with increasing age is more readily observed. The size of the means for Complex Mental Processes as compared with the size of the means for Fundamental Ability is not indicative, since these sizes represent the arbitrary numerical values of the respective scales by which the tests were graded.

The correlation of Fundamental Abilities with age is +.17 with a PE of +.025. This represents a low degree of correlation, yet when portrayed graphically, it reveals some significance. The correlation of Complex Mental Processes with age is +.95, with a PE of +.003. The Multiple correlation of age with (Fundamental Ability and Complex Mental Processes) is +.97 with a "Chance R" of +.056.

The comparison of the respective trends indicates, first of all, that these two batteries have tapped two different kinds or phases of ability. One of these we have defined as native endowment, Fundamental Ability, or mental growth considered as a biological function. The other is defined as a highly complex intellectual response to educational and cultural pressures; *i.e.* the development of Complex Mental Processes. From this point of view, mental growth at the

higher intellectual levels becomes less of a biological function and more of a response to selected environmental pressures.

It must be recalled that we are dealing with a chronological period ranging from eighteen through twenty-two years, hence what may be true of these particular ages, may not be true above or below them. Until now, the correlation between mental growth and age for this period has been assumed as zero, which would give us a plateau on the curve. By actual measurement, of the correlation, we have found it to be +.17, which gives a gradual incline upward toward maturity.

The trend of Complex Mental Processes ability, rises from eighteen to twenty-two years, with a slight drop of .11 at this last stage, possibly due to the small population at this point. Increasing age, lived under the stimulus of cultural and educational environment, has provided a more extensive orientation, a richer background from which all mental function products may be derived. The fruits of added years of experience are manifest in the trend toward greater ability to make highly complex responses with increasing age.

In order to present these correlations graphically, the slopes of the regression lines were determined by the formula:

$$M_{\frac{b}{a}} = \frac{\Sigma xy}{\Sigma x^2}$$

where $M$ represents the slope of the regression of (1) Fundamental Ability on age and (2) of Complex Mental Processes on age. $b$ represents the mean of either array of ability indices, and $a$ represents the array of ages. $x$ and $y$ represent deviations from the means of age and ability, respectively. The sum of these deviations is determined by the size of the populations in each group.

The regression of Fundamental Ability on age gave a coefficient of +.06 and the regression of Complex Mental Processes on age gave a coefficient of +.94. Further reduction of these values to sigma or standard units, facilitated their representation on the same graph. The formula in either case became:

$$M_{\frac{b}{a_s}} = \frac{M_{\frac{b}{a}}}{\sigma_b}$$

where $s$ means "standard units" or "sigma units" and $\sigma_b$ is the sigma of the array of either ability taken separately. This gave the slope of

the regression lines in terms of standard units; that for Fundamental Ability became $y = .15x$; and that for Complex Mental Processes, $y = .93x$.

The Standard Error for the Coefficient of Regression is obtained by the formula:

$$\sigma_{M_{\frac{b}{a}}} = \frac{\sigma_b}{\sigma_a}\sqrt{\frac{1 - r^2}{N}}$$

where $\sigma_a$ is the sigma of the age array; $r$ is the coefficient of correlation between ability and age, and $N$ the number of cases.

The Standard Error of the Coefficient of Regression, of Fundamental Ability on age is found to be +.013, and that for Complex Mental Processes on age is found to be +.012.

The values indicate first of all, the high degree of improbability that these two curves (1) and (2) can ever change sufficiently to move with the same degree of acceleration, within the range of eighteen to twenty-three years, thus furnishing added assurance, in terms of increased probability, to any conclusions justified by these curves.

The limits set upon the slope of the Fundamental Ability curve (2) by its standard of error, indicate that there is but very slight probability that the mental growth curvew ill reach a plateau during this period. In other words, there is exceedingly high probability, that this portion of the curve will always rise slightly, and furthermore, that its present location is the most probable.

In a similar manner, the probability that the Complex Mental Process curve (1) will ever reach a plateau during this period is even more remote, with the strong presumption that its present position is the most likely, for any random sampling at the College-Adult Level.

Diagram I is a graphic presentation of these results. Our problem is to reconcile these two different aspects of mental growth. It is evident from Section 2 of the curve, that the hitherto assumed plateau was relatively though not absolutely correct. That the curve is approaching a plateau at age seventeen is certain, but that this plateau lies beyond twenty-three is also certain. Fundamental Ability, native endowment, is still slightly on the increase through the twenty-second year.

Section 1 of the dual curve, resembles very much the curves obtained for test ability correlated with ages eight to fifteen on the National Intelligence Test; also that from eighteen months to six and

one half years on the Stutsman Test, and that from eleven to sixteen years on the Otis Advanced Intelligence Test.[1] The rate of its acceleration is in sharp contrast with that of Section 2. The relative height above Section 2 is insignificant, as this position was arbitrarily chosen with respect to other features of the dual curve. The relative slope of these two curves is the significant factor.

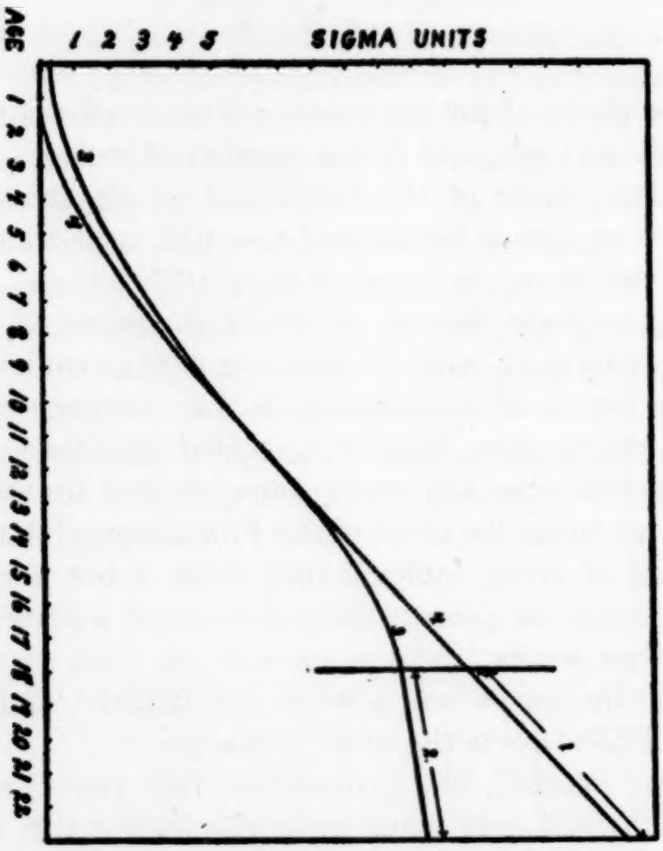



DIAGRAM I

**Curves showing the relative slope of the Regression Lines, from Ages eighteen to twenty-three.**

1. **Regression of Complex Mental Processes on Age, eighteen to twenty-three.**
2. **Regression of Fundamental Abilities on Age, eighteen to twenty-three.**
3. **Thurstone Mental Growth Curve from zero to seventeen years.**
4. **Theoretical Curve of Complex Mental Process Development, zero to seventeen.**

It would appear from the shape of Section 2, that the native endowment or mental growth regarded as a biological function, is fast approaching its maximum between eighteen and twenty-three years.

---

[1] Thurstone, L. L.: The Absolute Zero in Intelligence Measurement. *Psychological Review*, Vol. 35, No. 3, May, 1928, p. 193.

On the other hand, the intellectual development and cultural status of this same group, continues to rise proportionately to their continued subjection to educational and cultural curricula. It seems reasonable to characterize Section 2 as indicative of rather elementary mental behavior, and Section 1 as indicative of rather highly complex mental behavior. We shall probably avoid some confusion in discussion by omitting the doubtful inference to "intelligence."

A further characterization, which differentiates these two types of mental growth, might regard Section 2 as the evolution of the inherent qualities of the protoplasm—a simple behavioristic estimate of mental growth. The other curve, Section 1, represents no denial of the former, but reveals a supplementary aspect of mental function in personality development, wherein the cultural and educational influences of society are manifest. This curve shows a much higher correlation with chronological age. Since these two types of behavior are produced by the same biological organism—each as inevitable as the other—it would seem that every stimulus situation offered an opportunity for selective reaction or choice of mental function. We may react biologically, at the animal level, or we may react as intellectual personalities at a higher level. Perhaps the greater degree of employability, tends toward the higher rather than the lower level, at the College-Adult Level, especially if we have regard for the present value of correlation coefficients.

Our next problem is to relate our results to previous investigations of mental growth. On the basis of the results obtained in this investigation, we have attempted to interpolate both Sections 1 and 2 of our dual curve, backward to birth.

We have accepted the mental growth curve of Thurstone, from three to seventeen years, as most indicative of this period. Without modification of either curve, our Fundamental Ability curve, Section 2, appears to be a continuation of his findings. With some degree of confidence, the curve for mental growth, considered as a biological function, can now be drawn from birth to the twenty-third year. From the shape of Thurstone's curve (3), it is evident that the rate of this developmental process is not constant. It has a positive acceleration to approximately the eleventh year, where it reaches an inflection point. There negative acceleration sets in with steadily increasing swiftness until we reach the seventeenth or eighteenth year, at which point it appears to enter upon a period of constancy through the twenty-second year. This type of mental growth increases as

much or more in any single year, between seven and fourteen years, as it does during the whole five year period from eighteen through twenty-two.

We have noted the similarity between our Complex Mental Process curve, Section 1, and the curves derived for other ages on the National and Otis tests. We have not endeavored to fit our curve to any of these as yet, though we propose to make this attempt in the subsequent investigation of our results on the basis of sex. For the moment, we have placed our Complex Mental Process curve, Section 1 on the chart in an arbitrary manner, preserving only its relative slope, and theoretically extending it toward zero ability. We have designed that it intercept the Thurstone curve at the eleven year point. This would bring zero ability for this type of mental ability at about the four year age, if the curve were carried through.

If however, the Complex Mental Process curve represents the development of a certain type of mental behavior, we should prefer to begin our theoretical curve, possibly at about eighteen months, with the inception of language ability in the individual, even though we may not be able to measure it here.

Between its inception, and the sixth year, the Complex Mental Process curve (4) is almost equally accelerated with the Fundamental Abilitiy curve, though slightly below it. At the sixth year, the individual meets the educationally organized current of civilization, which is designed to shape and fashion his mental behavior. He is subjected to educational, social and cultural pressures, represented by the school curriculum, his contacts with teachers, and his contemporaries in intellectual competition. Hence at this point, a more rapid acceleration of his Complex Mental Process ability takes place, which overtakes his Fundamental Abilitiy curve about the eleventh year. From this point to his twenty-third year, our assumed Complex Mental Process curve follows the trend indicated by our investigation.

This may account somewhat for the pronounced change that takes place in the individual's mental functioning just before the onset of puberty. According to our interpretive curve, at this point, the individual's behavior is determined less and less by pronounced biological mental response, and more and more by an intellectual or cultural response. This is in no sense a denial that our solution of many problems, even at upper age levels, rises no higher than biological mentality, but it is to say that we can and do develop potentialities that make reaction on a complex mental level, just as possible,

and certainly more probable. What we speak of as "control," very often means to delay a "biological mental" response, and refer the solution of the problem to the higher level of Complex Mental Processes. It is an observed phenomenon, that a pronounced change in mental behavior characterizes the period from nine to twelve years. Possibly the wane of biological mental development and the continued increase of highly complex mental processes may partially account for this change. Our interpretive curves offer an interesting field in which to conduct further investigations.

Our final consideration is the prediction of behavior. Frequent reference to our interpretive curves, will help to clarify the discussion. These curves have established the normal course of mental development. They predict that fifty per cent of the individuals of a given age will be found above and fifty per cent below them, in similar samples of the population. This may be regarded as a general diagnosis of mental behavior up to the twenty-third year.

Prognosis is simply the theoretical projection of diagnosis, and our interpretive curves, offer the most probable direction for prognostic projection. For example, since fifty per cent of the populations investigated, are either above or below the curve at three years of age and at twenty-three alike, it seems reasonable to assume that the group which was above at three years will also be the group which is still above at twenty-three, simply because the curve of normal mental development lies between these two groups all the way. Any other assumption would deny the validity of the normal curve of mental growth. It is mathematically impossible for a point one sigma below the curve at three years, to reach a point one sigma above the curve at twenty-three years, without tracing a new curve in its progress, which would be entirely different, than the curve which we have found to represent normal mental development.

Diagnosis establishes the individual's relation to normality and measures his deviations from the normal mental growth curve, on a straight line, perpendicular with the chronological base line. Prognosis on the other hand, concerns the movement of this diagnostic deviation point, toward the right, in the direction in which the chronological base line extends, *but not parallel to it.* The Deviation point must move in a curve which is of the same general form as the normal mental growth curve. This is but another way of saying, what is widely recognized, that bright, normal and dull children develop mentally at approximately the same rate.

It is to be noted however that variability increases with age, hence positive and negative deviations increase slightly with age. This fact helps to authenticate any prognosis based on accurate diagnosis, because it gradually separates the distance between the normal curve and deviate curves, with increasing years, thus making their coincidence or crossing even more remote.

The shape of the Fundamental Abilitiy curve, Sections 2 and 3, is an ardent protest against any prognosis of adult behavior, which is based on an infancy curve. The most thorough diagnosis of the one to five year period of this curve, gives no indication whatever of its direction between ten and fifteen or between eighteen and twenty-three years. Furthermore, the cultural and intellectual development, represented by the Complex Mental Process curve, Sections 1 and 4, by which the individual will be more constantly evaluated as a social unit, are so incipient from one to five years, that prognosis with respect to this phase of mental development, in this period, is very doubtful. In other words, the degree of acceleration with which mental development takes place with increasing age, *i.e.*, the shape of the normal development curves, can not be determined by prognosis. Now that these curves have been determined by investigation, prognosis must follow the course they indicate.

Brotemarkle has shown in his "Mental Graph Types" that the prediction of human behavior can only be based upon accurate diagnosis of an individual, with respect to the two phases of mental development indicated by our interpretive curves. A superior biological mental basis is not absolute assurance of superior ability with respect to the development of Complex Mental Processes.

This is only to say that social non-conformity which reaches the acute stage of variability in criminal acts, is just as frequently the concomitant of normal biological mentality as it is of abnormal. Conversely the normal biological mental endowment is quite capable of attaining high degrees of culture and social conformity. Biological mental development tells only part of the story, and prognoses made on this basis alone are subject to a very large probable error, large enough even to invalidate the prognoses.

True prognosis must begin with accurate diagnosis, with respect to Fundamental Abilities and Complex Mental Processes, and must infer development in accordance with the curves. The fundamental endowment of the individual is basic, to be sure, but his ultimate value as a unit of society, is even more determined by the permutations and

combinations of Fundamental Abilities, *i.e.*, the development of his Complex Mental Processes.

The following conclusions are indicated:

1. Mental Behavior, as an index of mental growth at the College-Adult Level, should be analyzed into Fundamental Abilities and Complex Mental Processes.

2. The development of Fundamental Ability, regarded as a biological function, has not yet reached the plateau stage between eighteen and twenty-three years.

3. Complex Mental Processes are more largely due to environmental pressures than are Fundamental Abilities, hence they respond and increase more rapidly at the College-Adult Level.

4. Though we may not expect the Complex Mental Processes to increase indefinitely, there is no evidence of an approaching plateau at age twenty-three.

5. Fundamental Abilities develop as much, or more, in any single year between seven and fourteen, as they do during the entire five year period from eighteen through twenty-two, at the College-Adult Level.

6. There is a possibility that the probable proportionate increase of Complex Mental Process development, and the proportionate decrease of Fundamental Ability development, just prior to or about the age of puberty, may partially account for the changes in mental behavior observable at this period.

7. While children up to eleven years of age may be more inevitably committed to reactions at the level of biological mental development, they become increasingly committed to reactions at the higher complex intellectual level thereafter.

8. Predictions of behavior should be based upon the shape of the mental growth and behavior curves, which change conspicuously with increasing years.

I wish to express my gratitude to Dr. George Gailey Chambers and Dr. H. M. Lufkin, of the Department of Mathematics, of the University of Pennsylvania, for criticism and suggestions in the development of statistical procedures; to Dr. Samuel W. Fernberger of the Department of Psychology of the University of Pennsylvania for council and assistance in my investigation; to Dr. Robert A. Brotemarkle, under whose supervision the investigation was carried out, for materials and suggestions which greatly facilitated the work; and to the instructors and students, whose cooperation over a period of three years made the materials available.

## Bibliography

Brooks, Fowler D.: "Changes in Mental Traits with Age." Teachers College Publications, Columbia University, 1921.

Brotemarkle, Robert A.: College Student Personnel Problems. *Journal of Applied Psychology*, Vol. XI, No. 6, 1927.

Cunningham, Kenneth S.: "The Measurement of Early Levels of Intelligence." Teachers College Publications, Columbia University, 1927.

Fernberger, S. W.: Statistical and Non-statistical Treatment of Results. *Psychological Clinic*, Vol. XIV, Nos. 3–4, May–June, 1922.

Foran, T. G.: The Constancy of the Intelligence Quotient. *Educational Research Bulletin*, Vol. I, No. 10, December, 1926.

Garrett, H. E.: "Statistics in Psychology and Education."

MacPhail, Andrew H.: "The Intelligence of College Students."

Miller, Karl G.: The Competency of Fifty College Students. *The Psychological Clinic*, Vol. XIV, Nos. 1–2, March–April, 1922.

Rosenow, Curt: Analysis of Mental Functions. *Psychological Monographs*, Vol. XXIV, No. 5, June, 1917.

Terman, L. M.: "The Intelligence of School Children."

Thorndike, E. L.: "The Measurement of Intelligence."

Thurstone, L. L.: A Method of Scaling Psychological Tests. *Journal of Educational Psychology*, Vol. XVI, No. 7. Absolute Zero in Intelligence Measurement. *Psychological Review*, Vol. XXXV, No. 3. The Mental Growth Curve for Binet Tests. *Journal of Educational Psychology*, Vol. XX, No. 8.

Tilton, J. W.: "The Relation between Association and the Higher Mental Processes." Teachers College Publications, Columbia University, 1926.

# A MODIFICATION OF THE COMPUTATION OF THE MULTIPLE CORRELATION AND REGRESSION COEFFICIENTS BY THE TOLLEY AND EZEKIEL METHOD

AARON BAKST

Teachers College, Columbia University

The Tolley and Ezekiel method of handling the multiple correlation problem given in their paper[1] and further modified by Garrett[2] states the relation between the partial regression coefficients, $b$'s, in the form of a set of linear normal equations that can be solved simultaneously.

These equations are given in two forms: either in terms of the mean product sums, or in the terms of the correlation coefficients defined by the relations:

$$p_{12} = r_{12}\sigma_1\sigma_2, \quad r_{12} = \frac{\Sigma(x_1x_2)}{N\sigma_1\sigma_2}, \quad p_{12} = \frac{\Sigma(x_1x_2)}{N}$$

Rewriting Garrett's equations (3) (p. 42) we have:

$$\begin{aligned}
r_{12} &= A_{21}b_{12.34}\ldots + r_{23}A_{21}b_{13.24}\ldots + r_{24}A_{41}b_{14.23}\ldots + \cdots + r_{2n}A_{n1}b_{1n\cdot23}\ldots \\
r_{13} &= r_{23}A_{21}b_{12.34}\ldots + A_{21}b_{13.24}\ldots + r_{34}A_{41}b_{14.23}\ldots + \cdots + r_{3n}A_{n1}b_{1n\cdot23}\ldots \\
r_{14} &= r_{24}A_{21}b_{12.34}\ldots + r_{34}A_{21}b_{13.24}\ldots + A_{41}b_{14.23}\ldots + \cdots + r_{4n}A_{n1}b_{1n\cdot23}\ldots \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
r_{1n} &= r_{2n}A_{21}b_{12.34}\ldots + r_{3n}A_{21}b_{13.24}\ldots + r_{4n}A_{41}b_{14.23}\ldots + \cdots + A_{n1}b_{1n.23}\ldots
\end{aligned}$$

where the values of $A_{21}$, $A_{31}$, $A_{41}$, etc. are represented by $\sigma_2/\sigma_1$, $\sigma_3/\sigma_1$, $\sigma_4/\sigma_1$, etc., respectively.

---

[1] Tolley, H. R. and Ezekiel, M. J. B.: A Method of Handling Multiple Correlation Problems. *Journal of American Statistical Association*, Vol. XVIII, Dec., 1923, pp. 993–1003.

[2] Garrett, Henry E.: A Modification of Tolley and Ezekiel's Method of Handling Multiple Correlation Problems. *Journal of Educational Psychology*, Vol. XIX, Jan., 1928, pp. 45–49.

The value of $b_{12.34}\ldots$ can be written in determinantal form as:

$$b_{12.34\ldots n} = \frac{\begin{vmatrix} r_{12} & r_{23}A_{31} & r_{24}A_{41} & \ldots & r_{2n}A_{n1} \\ r_{13} & A_{31} & r_{34}A_{41} & \ldots & r_{3n}A_{n1} \\ r_{14} & r_{34}A_{31} & A_{41} & \ldots & r_{4n}A_{n1} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ r_{1n} & r_{3n}A_{31} & r_{4n}A_{41} & \ldots & A_{n1} \end{vmatrix}}{\begin{vmatrix} A_{21} & r_{23}A_{31} & r_{24}A_{41} & \ldots & r_{2n}A_{n1} \\ r_{23}A_{21} & A_{31} & r_{34}A_{41} & \ldots & r_{3n}A_{n1} \\ r_{24}A_{21} & r_{34}A_{31} & A_{41} & \ldots & r_{4n}A_{n1} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ r_{2n}A_{21} & r_{3n}A_{31} & r_{4n}A_{41} & & A_{n1} \end{vmatrix}}$$

The columns of both determinants of the numerator and determinator have common factors $A_{21}$, $A_{31}$, $A_{41}$, etc. that can be taken outside of the determinant. Moreover, the factors of the determinant of the numerator appear as factors in the determinant of the denominator. Therefore, the value of $b_{12.34\ldots n}$ can be rewritten as:

$$b_{12.34\ldots n} = \frac{\begin{vmatrix} r_{12} & r_{23} & r_{24} & \ldots & r_{2n} \\ r_{13} & 1 & r_{34} & \ldots & r_{3n} \\ r_{14} & r_{34} & 1 & \ldots & r_{4n} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ r_{1n} & r_{3n} & r_{4n} & \ldots & 1 \end{vmatrix}}{A_{21}\begin{vmatrix} 1 & r_{23} & r_{24} & \ldots & r_{2n} \\ r_{23} & 1 & r_{34} & \ldots & r_{3n} \\ r_{24} & r_{34} & 1 & \ldots & r_{4n} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ r_{2n} & r_{3n} & r_{4n} & \ldots & 1 \end{vmatrix}}$$

In general the value of $b_{1k.23\ldots n}$ can be expressed in this form, and noticing that in the determinant of the numerator we replace with

the column $r_{12}$, $r_{13}$, $r_{14}$, etc. the column in which the coefficients of $b_{1k.23\ldots n}$ appear, and moreover that transposing the $k$-th column in the position of the first column and the $k$-th row in the position of the first row of both determinants we do now change the value of $b_{1k.23\ldots n}$ and then replace the $k$-th column in the numerator in the same manner as in case of obtaining the expression of $b_{12.23\ldots n}$ the expression of $b_{1k.23\ldots n}$ is as follows:

$$b_{1k.23\ldots n} = \frac{\sigma_1}{\sigma_k} \frac{\begin{vmatrix} r_{1k} & r_{2k} & r_{3k} & \ldots & r_{kn} \\ r_{12} & 1 & r_{23} & \ldots & r_{2n} \\ r_{13} & r_{23} & 1 & \ldots & r_{3n} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ r_{1n} & r_{2n} & r_{3n} & \ldots & 1 \end{vmatrix}}{\begin{vmatrix} 1 & r_{2k} & r_{3k} & \ldots & r_{kn} \\ r_{2k} & 1 & r_{23} & \ldots & r_{2n} \\ r_{3k} & r_{23} & 1 & \ldots & r_{3n} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ r_{kn} & r_{2n} & r_{3n} & \ldots & 1 \end{vmatrix}} \tag{2}$$

The evaluation of these two determinants, which are identical, with the exception of their first columns can be done as follows:

Consider a determinant

$$D = \begin{vmatrix} a_1 & b_1 & c_1 & d_1 & \ldots & k_1 & l_1 \\ a_2 & b_2 & c_2 & d_2 & \ldots & k_2 & l_2 \\ a_3 & b_3 & c_3 & d_3 & \ldots & k_3 & l_3 \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot \\ a_n & b_n & c_n & d_n & \ldots & k_n & l_n \end{vmatrix}$$

Multiplying each term of every column by the first term of the column on the right, with its sign changed, and adding to it the product of the corresponding terms in the same row in the column on the right with the first term of the original column, we have:

$$(-1)^{n-1}b_1c_1d_1 \cdots k_1l_1D = \begin{vmatrix} 0 & 0 & \cdots & 0 & l_1 \\ -a_2b_1 + a_1b_2 & -b_2c_1 + b_1c_2 & \cdots & -k_2l_1 + k_1l_2 & l_2 \\ -a_3b_1 + a_1b_3 & -b_3c_1 + b_1c_3 & \cdots & -k_3l_1 + k_1l_3 & l_3 \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ -a_nb_1 + a_1b_n & -b_nc_1 + b_1c_n & & -k_nl_1 + k_1l_n & l_n \end{vmatrix}$$

and this by the expansion of the determinant by minors (taking them along the first row) is transformed into:

$$(-1)^{n-1}b_1c_1d_1 \cdots k_1l_1D = (-1)^{n-1}l_1 \begin{vmatrix} -a_2b_1 + a_1b_2 & -b_2c_1 + b_1c_2 & \cdots & -k_2l_1 + k_1l_2 \\ -a_3b_1 + a_1b_3 & -b_3c_1 + b_1c_3 & \cdots & -k_3l_1 + k_1l_3 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ -a_nb_1 + a_1b_n & -b_nc_1 + b_1c_n & \cdots & -k_nl_1 + k_1l_n \end{vmatrix}$$

or finally

$$D = \frac{1}{b_1c_1d_1 \ldots k_1} \begin{vmatrix} \begin{vmatrix} a_1b_1 \\ a_2b_2 \end{vmatrix} & \begin{vmatrix} b_1c_1 \\ b_2c_2 \end{vmatrix} & \begin{vmatrix} c_1d_1 \\ c_2d_2 \end{vmatrix} & \ldots & \begin{vmatrix} k_1l_1 \\ k_2l_2 \end{vmatrix} \\ \begin{vmatrix} a_1b_1 \\ a_3b_3 \end{vmatrix} & \begin{vmatrix} b_1c_1 \\ b_3c_3 \end{vmatrix} & \begin{vmatrix} c_1d_1 \\ c_3d_3 \end{vmatrix} & \ldots & \begin{vmatrix} k_1l_1 \\ k_3l_3 \end{vmatrix} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \begin{vmatrix} a_1b_1 \\ a_nb_n \end{vmatrix} & \begin{vmatrix} b_1c_1 \\ b_nc_n \end{vmatrix} & \begin{vmatrix} c_1d_1 \\ c_nd_n \end{vmatrix} & \ldots & \begin{vmatrix} k_1l_1 \\ k_nl_n \end{vmatrix} \end{vmatrix} \qquad (A)$$

Using this transformation, and noticing that the terms taken out in the fraction on the left of the determinant ($A$) will be equal for both determinants of (2) the expression for $b_{1k.23\ldots n}$ can be rewritten as follows:

$$b_{1k.23\ldots n} = \frac{\sigma_1}{\sigma_k} \frac{\begin{vmatrix} \begin{vmatrix} r_{1k} & r_{2k} \\ r_{12} & 1 \end{vmatrix} & \begin{vmatrix} r_{2k} & r_{3k} \\ 1 & r_{23} \end{vmatrix} & \cdots \\ \begin{vmatrix} r_{1k} & r_{2k} \\ r_{13} & r_{23} \end{vmatrix} & \begin{vmatrix} r_{2k} & r_{3k} \\ r_{23} & 1 \end{vmatrix} & \cdots \\ \cdots & \cdots & \end{vmatrix}}{\begin{vmatrix} \begin{vmatrix} 1 & r_{2k} \\ r_{2k} & 1 \end{vmatrix} & \begin{vmatrix} r_{2k} & r_{3k} \\ 1 & r_{23} \end{vmatrix} & \cdots \\ \begin{vmatrix} 1 & r_{2k} \\ r_{3k} & r_{23} \end{vmatrix} & \begin{vmatrix} r_{2k} & r_{3k} \\ r_{23} & 1 \end{vmatrix} & \cdots \\ \cdots & \cdots & \end{vmatrix}}$$

Continuing the process of transforming the determinants in the same manner the complete evaluation of the $b_{1k.23\ldots n}$ is concluded.

The advantage of this method of evaluation lies in the fact that in the case of each particular regression coefficient only one column is inserted in a new position in the determinant, and only one row and column are transposed in the determinants of the numerator and denominator. Therefore, the evaluation of any subsequent regression coefficients is simplified, for as it is seen from the nature of determinant ($A$) certain results are retained throughout the entire process of computing.

It can be readily seen that this process of computing offers another advantage, namely that any regression coefficients of orders less than "$n$" can be obtained during the computation of $b_{1k.23\ldots n}$.

Let us write down the set of equations predicting $b_{1k\cdot 23}$

$$\begin{aligned} r_{12} &= A_{21}b_{12.3k} + r_{23}A_{31}b_{13.2k} + r_{2k}A_{k1}b_{1k.23} \\ r_{13} &= r_{23}A_{21}b_{12.3k} + A_{31}b_{13.2k} + r_{2k}A_{k1}b_{1k.23} \\ r_{1k} &= r_{2k}A_{21}b_{12.3k} + r_{2k}A_{31}b_{13.2k} + A_{k1}b_{1k.23} \end{aligned}$$

The reader can readily construct the value of $b_{1k.23}$ in the form of determinants, it is (following the process described above and transposing rows and columns)

$$C_{11.23}\frac{\sigma_1}{\sigma_k}\frac{\begin{vmatrix} r_{1k} & r_{2k} & r_{3k} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{vmatrix}}{\begin{vmatrix} 1 & r_{2k} & r_{3k} \\ r_{2k} & 1 & r_{23} \\ r_{3k} & r_{23} & 1 \end{vmatrix}}$$

It may be observed that when solving the original equations for $b_{1k.23\ldots n}$ the values of $b_{1k.2}$, $b_{1k.23}$, $b_{1k.234}$, etc., may be computed.

The formula for the coefficient of multiple correlation, $R$, as given by Garrett (p. 48, formula 7) is

$$R^2_{1(23\ldots n)} = \frac{b_{12.34\ldots n}r_{12}\sigma_2 + b_{13.24\ldots n}r_{13}\sigma_3 + \cdots + b_{1n.23\cdots(n-1)}r_{1n}\sigma_n}{\sigma_1 b}$$

substituting in this formula the values of $b_{1k.23\ldots n}$ from formula (2) we have

$$R^2_{1(23\ldots n)} = b_{12.34\ldots n}r_{12}A_{21} + b_{13.24\ldots n}r_{13}A_{31} + \cdots + b_{1n.23\ldots(n-1)}r_{1n}A_{n1}$$

or in other form

$$R^2_{1(23\ldots n)} = \sum_{k=2}^{n} r_{1k}\frac{\begin{vmatrix} r_{1k} & r_{2k} & r_{3k} & \ldots & r_{kn} \\ r_{12} & 1 & r_{23} & \ldots & r_{2n} \\ r_{13} & r_{23} & 1 & \ldots & r_{3n} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ r_{1n} & r_{2n} & r_{3n} & \ldots & 1 \end{vmatrix}}{\begin{vmatrix} 1 & r_{2k} & r_{3k} & \ldots & r_{kn} \\ r_{2k} & 1 & r_{23} & \ldots & r_{2n} \\ r_{3k} & r_{23} & 1 & \ldots & r_{3n} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ r_{kn} & r_{2n} & r_{3n} & \ldots & 1 \end{vmatrix}}$$

which again enables the computation of all the $R$'s of orders less than "$n$" as well.*

Similarly, this process allows the calculation of the standard errors of estimate of all orders up to and including "$n$," as given by the formula

$$\sigma^2_{1.23\ldots n} = \sigma_1\sqrt{1 - R^2_{1(23\ldots n)}}$$

*The row and the column where $r_{kk}$ appears must be stricken out in both determinants. The same applies to equation (2).

# THE EFFECT OF THE 6-22-44-22-6 NORMAL CURVE SYSTEM ON FAILURES AND GRADE VALUES

J. DE WITT DAVIS

University of Oregon

There appeared recently a study[1] setting up the 6–22–44–22–6 per cent procedure for a five point grading scheme. It has distinct theoretic advantage over those recommended by others, among them Dearborn[2] and Rugg.[3] Any use of the normal curve as a basis for grade distribution arouses both strong opposition particularly from those who do not know how to apply it, and equally strong support from those who have given it the most study and most careful application. The growing employment of one or another modification of it prompts this analytical study.

The purpose here is not to show how either the small or large class as such is affected. The task assumed is twofold, namely:

1. Aside from other factors of test validity, and reliability, and also aside from other causes for leaving school, and aside from entrances other than as a freshman, what is the effect of a strict application of the 6–22–44–22–6 system as a basis for college grades from the standpoint of college failure?

2. And further, what is the comparative value or relative worth of grades from class to class when this system is applied?

For convenient thinking let us assume that after the weeding out of Freshman Week there remain a group of one thousand students for college work in a course which continues consecutively throughout the four college years with three terms in each year. No others join this purely hypothetical group, none leave it except by graduation as seniors, or because of being failed by the system. How many survive at the various levels?

The adopted system requires a brief explanation. By nature it forces each new term group into a normal distribution. It automati-

---

[1] Eells, Walter Crosby: An Improvement in the Theoretical Basis of the Five Point Grading System Based on the Normal Probability Curve. *Journal of Educational Psychology*, Vol. XXI, No. 2, Feb., 1930, pp. 128–136.

[2] Dearborn, W. F.: School and University Grades. *University of Wisconsin Bulletin*, No. 368, Madison, Wisconsin, 1910.

[3] Rugg, Harold O.: "Statistical Methods Applied to Education." Houghton Mifflin Co., Boston, 1917, pp. 216–219.

cally fails six per cent, gives twenty-two per cent a grade of IV, forty-four per cent a grade of III, twenty-two per cent a grade of II and six per cent a grade of I. Since the group is forced into normalcy it may be assumed that practically one hundred per cent is included within $\pm 3\sigma$ range. AI then begins at approximately $1.6\sigma$ and takes the balance up; AII begins at a $+.6\sigma$ and extends up to the former; AIII begins at a $-.6\sigma$ and goes up to the lowest II; AIV begins at a $-1.6\sigma$ and extends to the lowest III; while AV takes the balance, $-1.6\sigma$ down.

A table constructed on the basis of this system shows clearly the elimination that takes place each of the twelve school terms when there are three terms per year for four years.

TABLE OF FAILURES PER SCHOOL TERM

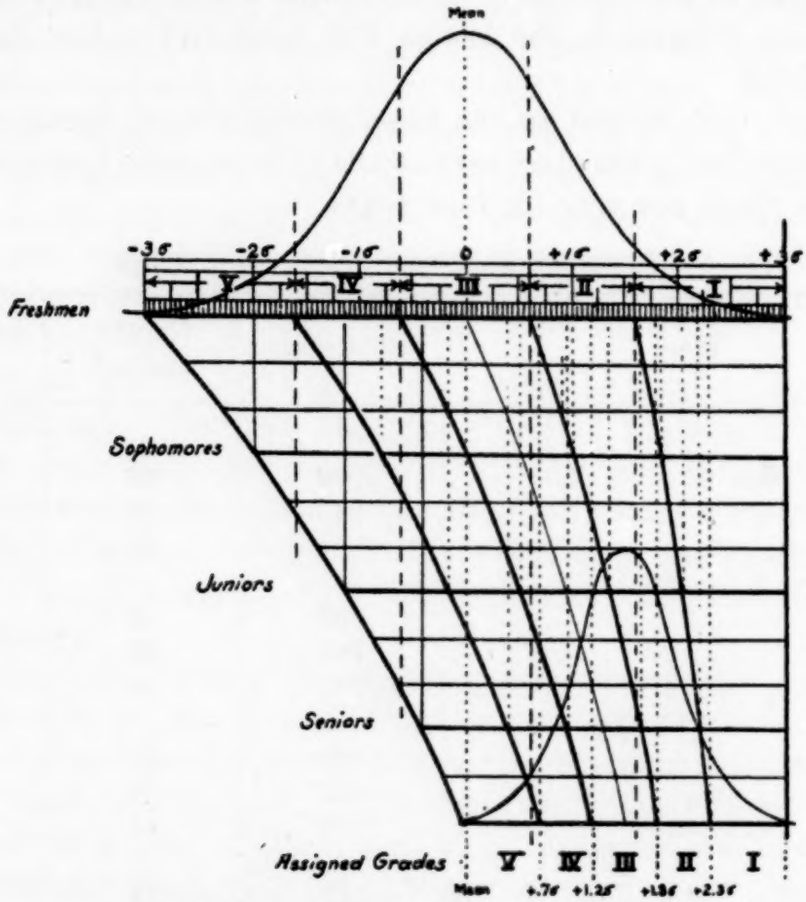
| Term | Entering | Number failing* | Accumulated failures |
|---|---|---|---|
| Freshman | | | |
| 1st | 1000 | 60 | 60 |
| 2nd | 940 | 56 | 116 |
| 3rd | 884 | 52 | 168 |
| Sophomore | | | |
| 1st | 832 | 49 | 217 |
| 2nd | 783 | 46 | 263 |
| 3rd | 737 | 44 | 307 |
| Junior | | | |
| 1st | 693 | 41 | 348 |
| 2nd | 652 | 39 | 387 |
| 3rd | 613 | 36 | 423 |
| Senior | | | |
| 1st | 577 | 34 | 457 |
| 2nd | 543 | 32 | 489 |
| 3rd | 511 | 30 | 519 |
| Total graduates | 481 | 519 | 519 |

* Fractions are interpreted in favor of the students, *e.g.* 29½ = 30 students.

Assuming that the measures employed are perfectly valid and perfectly reliable, the poorer half have all been eliminated. This may or may not be desirable, a Platonic question might easily arise; if desirable, why not eliminate them at first and save the waste in time and expense both to the student and to the state? The answer might then come back that the measures really aren't perfect, besides, it is held by some to be essential that a student try out for a special

ability in some given lines. These are only implications aroused by the practice and therefore are only parenthetical to the question.

The table above answers the first query, namely, a strict application of the 6–22–44–22–6 plan gradually eliminates by failing a total fifty-one per cent of the admitted freshmen, sixty at the first term level and thirty at the last term of senior work.




6-22-44-22-6 Normal Curve Application

As to grades, what is their relative value from term to term, when assigned in this manner? By the use the geometric progression formula[1] the nomograph following was constructed. Since the upper end of the distribution was not affected by failure the left or lower end was shortened. The border lines of the grade spreads as given above were determined by the use of Holzinger's Tables.[2] Notice

[1] $t_n = ar^{n-1}$ when $t$ = the term wanted, $n$ the number of the term in the series, $a$ the first term of the series, and $r$ the ratio.

[2] Holzinger, Karl J.: "Statistical Tables for Students in Education and Psychology." The University of Chicago Press, 1925.

on the nomograph that the mean at each level, by conforming to the normal distribution is caused to migrate to the right during the twelve terms a total distance of $1.5\sigma$ on the original base line. It is from this new mean always that the fixed distribution is made each term. A brief study of the graph makes the results evident.

A table of compared grades for the four classes as pictured by this graph will be sufficient.

FIRST TERM OF EACH YEAR
Comparison of Grade Values

| Freshmen grades | Sophomores equivalents | Juniors equivalents | Seniors equivalents |
|---|---|---|---|
| I | All of Is<br>Upper IIs | All of Is<br>Most of IIs | All of Is<br>All of IIs<br>A few IIIs |
| II | Lower IIs<br>About one-half of IIIs | Lower of IIs<br>Nearly all of IIIs | Balance of IIIs<br>Over one-half of IVs |
| III | Balance of IIIs<br>Most of IVs | Balance of IIIs<br>All of IVs<br>Over one-half of Vs | Balance of IVs<br>All of Vs |
| IV | Balance of IVs<br>Most of Vs | Balance of Vs | |
| V | Balance of Vs | | |

Similar tables of comparison can be worked out by the interested reader using the grade of any other level as a criterion. It is evident that the farther removed from each other by school terms the grades may be the less comparable they become.

## CONCLUSION

1. A strict application of the 6-22-44-22-6 normal curve procedure as a basis for grades in a twelve term course would eliminate practically fifty-one per cent of the matriculated group.

2. This strict practice changes the values of grades from term to term so that a grade of I or II, etc. has little directly comparable meaning unless associated with a particular term level.

Much has been written about predicting grades from those made in former work: this study seems to indicate, though the measures were assumed to be perfectly valid and reliable, that predictability would yet be low due to this lack of comparability especially so when term grades of seniors are predicted say from freshman or sophomore grades. Readers may desire another system to provide for apparent skewness, at the same time to employ the basic technique of the normal curve and yet to keep the value of grades equal but that is another study. Others will recall this-and-that college which still has the two semester year division and will at once start comparisons with the twelve term schools. Many such implications arise calling for further consideration.